ABSTRACT

Title of Dissertation: EXPERIMENTAL, COMPUTATIONAL AND, THEORETICAL ANALYSIS OF THE GROEL/GROES CATALYTIC CYCLE

Nicholas Crane Corsepius
Doctor of Philosophy, 2014

Dissertation Directed By: Professor George H. Lorimer
Department of Chemistry and Biochemistry
Biochemistry Graduate Program

The GroEL/GroES system is among the best characterized molecular chaperonins, having been the focus of numerous studies since its discovery just over 40 years ago. There are several aspects of the chaperonin's catalytic cycle that have been thoroughly established, such as the positive intra-ring and negative inter-ring cooperativity that characterizes the binding of ATP, or the concerted domain movements within a ring that constitute the T → R allosteric transition. However, there are still several aspects of the GroEL/GroES catalytic cycle that are still not well understood, and the focus of this work is to address several of them. The results presented throughout this dissertation differ, sometimes drastically, from conventional wisdom, providing a new framework for the investigation of GroEL/GroES catalysis.

The work presented here contains several new insights into the GroEL/GroES catalytic cycle. The complicated nature of the interaction between GroEL and its substrate proteins (SPs) has previously prevented quantitative analysis of the role that SP binding plays in the cycle. The problem is circumvented in this work by application of an SP surrogate that simultaneously mimics the effects of bound SP and allows for the quantification of allosteric states. Development

of the system leads to the first ever quantification of the role of SP in the catalytic cycling of GroEL.

In addition to cycle features that had not been previously addressed, there are several parts of the catalytic cycle that have simply been overlooked throughout the years or unduly simplified in previous analysis. The hydrolysis of ATP within a ring and the release of ADP from a ring are often characterized as concerted events. Both steps are crucial to the cycling of the GroEL/GroES system, and are thoroughly investigated in this work. Thorough examination reveals that neither step is well-characterized as a concerted process. Pre-steady state kinetic analysis demonstrates that hydrolysis of ATP occurs stochastically, each subunit hydrolyzing ATP independently and randomly.

Pre-steady state kinetic analysis also reveals that ADP release is a multi-step process. Insight into the steps preceding/ accompanying ADP release are found in molecular dynamics (MD) simulations of the R → T allosteric relaxation. These simulations demonstrate the allosteric relaxation proceeds through a highly asymmetric series of states. Furthermore, the relaxation path taken by subunits can be described as exhibiting dynamic negative-cooperativity. The analysis implicates the allosteric relaxation as the trigger that allows the GroEL/GroES system to effectively respond to environmental stimuli.

The last topic addressed concerns the role of the GroEL/$GroES_2$ symmetric complex and the mechanism governing its cycling. It has long been assumed that a ring must hydrolyze all seven ATP before it can discharge GroES. This assumption is challenged by new insight into the stochastic mechanism of ATP hydrolysis. The analysis presented in this paper undermines the idea of all-or-none cycling and builds a quantitative relationship between the nucleotide distribution in a complex and its propensity to dissociate.

Experimental, Computational and, Theoretical Analysis of the GroEL/GroES Catalytic Cycle

Nicholas Crane Corsepius

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:

Professor George H. Lorimer, Chair
Professor Dorothy Beckett
Professor Edward Eisenstein
Professor David Fushman
Professor Devarajan Thirumalai

ACKNOWLEDGEMENTS

All I can say is that I am a very lucky man to be able to live that life that I am living, and I will do my best to make the most of this blessing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADP – Adenosine diphosphate

Apo – Unliganded protein

ATP – Adenosine triphosphate

BoS – Solution to the breakage of symmetry problem (Chapter 4)

D398A – GroEL mutant that hydrolyzes ATP at 0.1% of the wild-type rate

DTT – Dithiothreitol

EL – Abbreviation for GroEL

EL242C – GroEL mutant with covalently label-able cysteine in SP binding site

EL242C-TMR – EL242C with TMR covalently bound to residue 242C

ES – Abbreviation for GroES

F5M – Fluoresceine-5-maleimide

FRET – Forster resonance energy transfer

kDa – kilo-Dalton

KNF – Sequential cooperativity model

MD – Molecular dynamics

MDCC - N-[2-(1-maleimidyl)ethyl]-7-(diethylaminio)-coumarin-3-carboxamide. A fluorescent dye

MWC – Concerted cooperativity model

NCM – Nested cooperativity model

ODE – Ordinary differential equation

PBP – Phosphate binding protein

Pi – Inorganic phosphate

SOP – Self-organized polymer model

SP – Substrate protein

SVD – Singular value decomposition

T5M – Tetramethylrhodamine-5-maleimide

TMR – Tetramethylrhodamine

TMR_2-DTT – Compound with two T5M bound to one DTT molecule

Chapter 0

Introduction and Specific Research Aims

0.1 BACKGROUND

In 1973, in his seminal work "Principles that Govern the Folding of Protein Chains", Christian Anfinsen laid the foundation of our current understanding of protein folding, which includes the idea that a protein's primary structure completely dictates its active native confirmation (1). In the decade that followed, biochemists unraveled a series of mysteries that led them to several exceptions to this dogma (2,3). In 1983, the term "molecular chaperone" was first used to describe the class of enzymes that catalyze the folding of other proteins (4). GroEL, along with its co-enzyme GroES, were among the first molecular chaperones isolated (5), and have been the subject of numerous published studies since that time.

In the catalytic cycle of GroEL/GroES, the folding of substrate proteins (SPs) is coupled to the hydrolysis of ATP. GroEL consists of fourteen 57kDa subunits, arranged in two back-to-back heptameric rings (6,7). A GroEL subunit contains three distinct domains: the apical, intermediate and equatorial domains. The equatorial domain houses the nucleotide-binding pocket, while the site for SP and GroES binding is found between helices H and I in the apical domain (8). The SP binding sites, primarily composed of hydrophobic residues, line the inner wall of the heptameric ring's central cavity, and bind the exposed hydrophobic regions of the misfolded SP. The "mobile loops" of GroES displace the SP from this site (9,10). In the chaperonin cycle, SP first binds to, and then becomes encapsulated inside GroEL's barrel-like structure. After a short time (a few seconds at 37°C) the SP is released, whether folded or not.

The chaperonin cycle involves a series of allosteric transitions within a heptameric ring (11). In the absence of ATP, the ring adopts the T state, which has a high affinity for SP, low affinity for nucleotide and, cannot bind GroES. The binding of ATP triggers a transition to the R state, which is accompanied by a rotation of the apical domains, the expansion of the central cavity and, the separation of the peptide binding sites (12). The R state has a high affinity for nucleotide, low

affinity for SP and, is the acceptor state for GroES. ATP binding is positively cooperative between subunits within a ring and negatively cooperative between rings within a 14mer (13-15).

The allosteric transition to the R state allows for the binding of GroES. The binding of GroES displaces bound SP and leads to a second allosteric transition from the R state to the R" state (9, 16). The transition is characterized by further apical domain movements, which yield an expanded central cavity that now accommodates the displaced SP. GroES acts as a lid on the cavity, fully encapsulating SP in the GroEL/GroES complex. The hydrolysis of ATP within a ring sets the timer for the lifetime of the GroEL/GroES complex (17, 18). Once the ATP has been hydrolyzed within complexed, or *cis*, ring, the binding of ATP and GroES to the adjacent, or *trans*, ring leads to the dissociation of GroES from the *cis* ring (19), and the ejection of the newly folded SP back into the bulk (20, 21).

In the presence of Mg^{2+} and K^{+}, GroEL hydrolyzes ATP to ADP and inorganic phosphate (22). In the absence of SP, the release of ADP is the rate-limiting step in the catalytic cycle (23). The addition of SP increases the rate of ATP consumption by GroEL by accelerating ADP/ATP exchange (24). A single SP binds to the apical domains of two or more GroEL subunits introducing a load on the ring, in the form of non-covalent inter-subunit cross-links, that stabilizes the T state (25, 26). This shifts the allosteric equilibrium in favor of the T state whose weak affinity for nucleotide leads to the dissociation of the product, ADP, effectively increasing the rate of ATP consumption.

0.2 OVERVIEW

This dissertation contains a series of experimental, computational and theoretical studies of the catalytic cycle of GroEL/GroES which have provided new insight into some of the most well-studied aspects of the chaperonin cycle. These studies have advanced our understanding of some of the catalytic steps that have often been overlooked or over-simplified by assumption. The

results provide surprising insight into the mechanism of hydrolysis and ADP release. The concluding catalytic model supports years of work on the GroEL/GroES system, valid under a variety of varying conditions.

The first chapter is an analysis of the T → R allosteric equilibrium of GroEL, using tetramethylrhodamine dimers. The allosteric equilibrium is probed using the dye dimers as an SP surrogate, which also spectroscopically reports distance information. The unique approach used in the study allows for the first quantitation of the impact SP has on the allosteric equilibrium. It also conclusively demonstrates the ability of the allosteric transition to perform work on bound SPs, lending support to the iterative annealing model of GroEL/GroES folding catalysis (27).

The second chapter is a detailed computational analysis of the pre-steady state kinetics of inorganic phosphate release (28, 29). The analysis compares a series of closely related catalytic models to determine which best describes the experimental data. This establishes that the hydrolysis of ATP is well described by a single step, a result used in the subsequent chapter to arrive at a mechanism of hydrolysis. The investigation also demonstrates that the release of ADP must be a multi-step process, a finding currently absent from the literature. Molecular dynamics simulations presented in the fifth chapter shed light on the mechanistic nature of this multi-step process.

The third chapter is a detailed computational study of the pre-steady state kinetics of the formation of the GroEL/GroES complex (28, 29). The results of this analysis lead to the conclusion that ATP hydrolysis is a multi-step process; a result that appears to contradict the findings of the study in Chapter 2. These two seemingly contradictory findings can be resolved only if the mechanism of hydrolysis is stochastic. A stochastic hydrolysis mechanism conflicts with the widely assumed concerted mechanism. The chapter concludes with a theoretical

comparison of the two hydrolysis models using single molecule simulations, giving insight into why a stochastic mechanism may be evolutionarily preferable.

The fourth chapter reformulates the "timer" mechanism linking ATP hydrolysis to the release of GroES and SP, in light of the results of the previous chapter. The link between the configuration of hydrolyzed nucleotides and GroEL/$GroES_2$ complex symmetry is formulated and fit to experimental results. The analysis concludes that hydrolysis of as few as five out of the seven ATPs within a ring is sufficient to cause GroES dissociation, and the actual number depends on the number of ATPs in the adjacent ring. The mechanism of the breakage of complex symmetry is especially important because it determines the lifetime of the cavity under physiological conditions.

The final chapter contains the results of a series of molecular dynamics (MD) simulations performed on the T → R and R → T allosteric transition (30). These simulations utilize newly published structural data on the nucleotide bound state. The results of these simulations are largely consistent with previous studies, illustrating a series of rigid domain movements that make up the transition. However, surprising ring asymmetry is observed in the R → T transition. Further inspection shows that the subunits interact in a dynamically cooperative fashion, forming stable docked pairs that reluctantly relax to the final symmetric structure. These results support one of the conclusions of Chapter 2, suggesting the potential mechanistic nature of the multi-step release process observed.

0.3 REFERENCES

1. Anfinsen, C. B. "Principles That Govern the Folding of Protein Chains." *Science* 181, no. 4096 (1973): 223-30.

2. Lorimer, G. H. "A Personal Account of Chaperonin History." *Plant Physiology* 125, no. 1 (2001): 38-41.

3. Horwich, A. L., G. W. Farr and W. A. Fenton. "Groel-Groes-Mediated Protein Folding." *Chem Rev* 106, no. 5 (2006): 1917-30.

4. Ellis, J. "Proteins as Molecular Chaperones." *Nature* 328, no. 6129 (1987): 378-9.

5. Hohn, T; Hohn, A; Engel, M; Wurtz; Smith, PR;. "Isolation and Characterization of the Host Protein Groe Involved in Bacteriophage Lambda Assembly." *Journal of Molecular Biology* 129, (1979): 359-373.

6. Hemmingsen, S. M., C. Woolford, S. M. van der Vies, K. Tilly, D. T. Dennis, C. P. Georgopoulos, R. W. Hendrix and R. J. Ellis. "Homologous Plant and Bacterial Proteins Chaperone Oligomeric Protein Assembly." *Nature* 333, no. 6171 (1988): 330-4.

7. Braig, K., Z. Otwinowski, R. Hegde, D. C. Boisvert, A. Joachimiak, A. L. Horwich and P. B. Sigler. "The Crystal Structure of the Bacterial Chaperonin Groel at 2.8 A." *Nature* 371, no. 6498 (1994): 578-86.

8. Chen, L. and P. B. Sigler. "The Crystal Structure of a Groel/Peptide Complex: Plasticity as a Basis for Substrate Diversity." *Cell* 99, no. 7 (1999): 757-68.

9. Landry, S. J., J. Zeilstra-Ryalls, O. Fayet, C. Georgopoulos and L. M. Gierasch. "Characterization of a Functionally Important Mobile Domain of Groes." *Nature* 364, no. 6434 (1993): 255-8.

10. Xu, Z., A. L. Horwich and P. B. Sigler. "The Crystal Structure of the Asymmetric Groel-Groes-(Adp)7 Chaperonin Complex." *Nature* 388, no. 6644 (1997): 741-50.

11. Horovitz, A., Y. Fridmann, G. Kafri and O. Yifrach. "Review: Allostery in Chaperonins." *J Struct Biol* 135, no. 2 (2001): 104-14.

12. Ranson, N. A., G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich and H. R. Saibil. "Atp-Bound States of Groel Captured by Cryo-Electron Microscopy." *Cell* 107, no. 7 (2001): 869-79.

13. Gray, T. E. and A. R. Fersht. "Cooperativity in Atp Hydrolysis by Groel Is Increased by Groes." *FEBS Lett* 292, no. 1-2 (1991): 254-8.

14. Yifrach, O. and A. Horovitz. "Two Lines of Allosteric Communication in the Oligomeric Chaperonin Groel Are Revealed by the Single Mutation Arg196-->Ala." *J Mol Biol* 243, no. 3 (1994): 397-401.

15. Yifrach, Ofer and A. Horovitz. "Nested Cooperativity in the Atpase Activity of the Oligomeric Chaperonin Groel." *Biochemistry* 34, no. 16 (1995): 5303.

16. Roseman, A. M., S. Chen, H. White, K. Braig and H. R. Saibil. "The Chaperonin Atpase Cycle: Mechanism of Allosteric Switching and Movements of Substrate-Binding Domains in Groel." *Cell* 87, no. 2 (1996): 241-51.

17. Lorimer, G. "Protein Folding. Folding with a Two-Stroke Motor." *Nature* 388, no. 6644 (1997): 720-1, 723.

18. Xu, Z. and P. B. Sigler. "Groel/Groes: Structure and Function of a Two-Stroke Folding Machine." *J Struct Biol* 124, no. 2-3 (1998): 129-41.

19. Horowitz, G. H. Lorimer and Ybarra. "GroES in the Asymmetric GroEL/ GroES Complex Exchanges Via an Associative Mechanism." *PNAS* 96, no. 6 (1999): 2682.

20. Weissman, J. S., C. M. Hohl, O. Kovalenko, Y. Kashi, S. Chen, K. Braig, H. R. Saibil, W. A. Fenton and A. L. Horwich. "Mechanism of Groel Action: Productive Release of Polypeptide from a Sequestered Position under Groes." *Cell* 83, no. 4 (1995): 577-87.

21. Todd, M. J., P. V. Viitanen and G. H. Lorimer. "Dynamics of the Chaperonin Atpase Cycle: Implications for Facilitated Protein Folding." *Science* 265, no. 5172 (1994): 659-66.

22. Viitanen, P. V., T. H. Lubben, J. Reed, P. Goloubinoff, D. P. O'Keefe and G. H. Lorimer. "Chaperonin-Facilitated Refolding of Ribulosebisphosphate Carboxylase and Atp Hydrolysis by Chaperonin 60 (Groel) Are K+ Dependent." *Biochemistry* 29, no. 24 (1990): 5665-71.

23. Grason, J. P., J. S. Gresham, L. Widjaja, S. C. Wehri and G. H. Lorimer. "Setting the Chaperonin Timer: The Effects of K+ and Substrate Protein on Atp Hydrolysis." *Proceedings of the National Academy of Sciences* 105, no. 45 (2008): 17334-17338.

24. Grason, J. P., J. S. Gresham and G. H. Lorimer. "Setting the Chaperonin Timer: A Two-Stroke, Two-Speed, Protein Machine." *Proc Natl Acad Sci U S A* 105, no. 45 (2008): 17339-44.

25. Stan, G., B. R. Brooks, G. H. Lorimer and D. Thirumalai. "Identifying Natural Substrates for Chaperonins Using a Sequence-Based Approach." *Protein Sci* 14, no. 1 (2005): 193-201.

26. Motojima, Charu Chaudhry, W. A. Fenton, G. W. Farr and A. L. Horwich. "Substrate Polypeptide Presents a Load on the Apical Domains of the Chaperonin GroEL" *PNAS* 101, no. 42 (2002): 15005.

27. Todd, M. J., G. H. Lorimer and D. Thirumalai. "Chaperonin-Facilitated Protein Folding: Optimization of Rate and Yield by an Iterative Annealing Mechanism." *Proc Natl Acad Sci U S A* 93, no. 9 (1996): 4030-5.

28. Ye, X. and G. H. Lorimer. "Substrate Protein Switches Groe Chaperonins from Asymmetric to Symmetric Cycling by Catalyzing Nucleotide Exchange." *Proc Natl Acad Sci U S A* 110, no. 46 (2013): E4289-97.

29. Yang, D., X. Ye and G. H. Lorimer. "Symmetric Groel:Groes2 Complexes Are the Protein-Folding Functional Form of the Chaperonin Nanomachine." *Proc Natl Acad Sci U S A* 110, no. 46 (2013): E4298-305.

30. Hyeon, G. H. Lorimer and D. Thirumalai. "Dynamics of Allosteric Transitions in Groel." *PNAS* 103, no. 50 (2006): 18939.

Chapter 1

Probing GroEL's Allosteric Equilibrium with Tetramethylrhodamine Dimers

1.1 INTRODUCTION

The GroEL/GroES nano-machine facilitates the folding of a large number of diverse substrate proteins (SPs), converting them from one of many misfolded states to unique native states, which are no longer recognized by GroEL. This machine is ultimately powered by the binding and hydrolysis of ATP, driving GroEL through a series of allosteric states (1). Two views of GroEL function have emerged: one as a passive anti-aggregation chamber and the other as an active folding device, performing work on SP. Our experiments support the latter view, demonstrating that GroEL has the ability to perform work on its SPs via domain movements that accompany the allosteric transitions.

1.1.1 Allostery Background

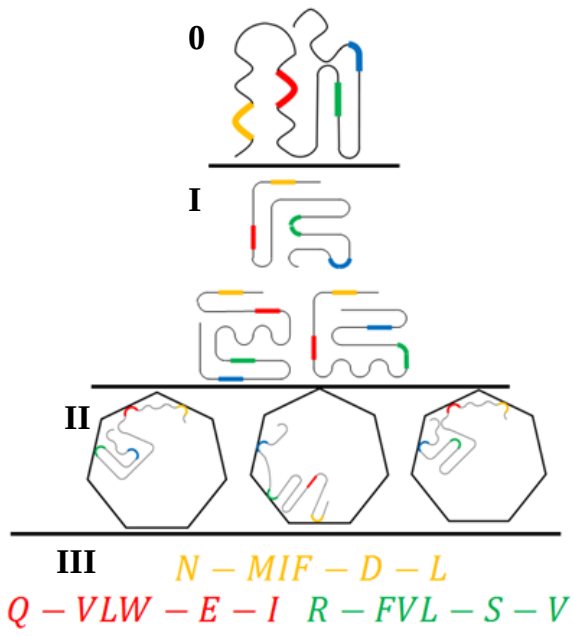
The chaperonin protein GroEL consists of fourteen 57kDa subunits, arranged in two back-to-back heptameric rings (3). A GroEL subunit contains three distinct domains: the apical, intermediate and equatorial domains. The equatorial domain houses the nucleotide-binding pocket, while the site for SP and GroES binding is found between helices H and I in the apical domain (4). The SP binding sites, primarily composed of hydrophobic residues, line the inner wall of the heptameric ring's central cavity, and bind the exposed hydrophobic regions of the misfolded SP. The "mobile loops" of GroES displace the SP from this site (5,6). In the chaperonin cycle, SP first binds to, and then becomes encapsulated inside GroEL's barrel-like structure. After a short time (a few seconds at 37°C) the SP is released, whether folded or not.

The chaperonin cycle involves a series of allosteric transitions within a heptameric ring (7). In the absence of ATP, the ring adopts the T state, which has a high affinity for SP, low affinity for nucleotide and, cannot bind GroES. The binding of ATP triggers a transition to the R state, which is accompanied by a rotation of the apical domains, the expansion of the central cavity and, the separation of the peptide binding sites (8). The R state has a high affinity for nucleotide, low affinity

for SP and, is the acceptor state for GroES. ATP binding is positively cooperative between subunits within a ring and negatively cooperative between rings within a 14mer. In the absence of GroES, GroEL can adopt three distinct allosteric states: TT, TR and RR.

In the presence of Mg^{2+} and K^{+}, GroEL hydrolyzes ATP to ADP and inorganic phosphate (9). In the absence of SP, the release of ADP is the rate-limiting step in the catalytic cycle (10). The addition of SP increases the rate of ATP consumption by GroEL by accelerating ADP/ATP exchange. A single SP binds to the apical domains of two or more GroEL subunits introducing a load on the ring, in the form of non-covalent inter-subunit cross-links, that stabilize the T state. This shifts the allosteric equilibrium in favor of the T state whose weak affinity for nucleotide leads to the dissociation of the product, ADP, effectively increasing the rate of ATP consumption (11).

Figure 1.1 – The Substrate Protein Binding Problem



A schematic illustration of the heterogeneous nature of SP binding. Level 0: SP's native confirmation. Level I: Every SP has multiple misfolded states. Level II: Each misfolded state can bind to GroEL with different topological arrangements. Level III: Each SP binding contact will differ because GroEL binds the general sequence P-HHH-P-H.

1.1.2 SP Binding Problem

Many GroEL SPs contain multiple "GroES-like" motifs of the generic sequence P-HHH-P-H, where P and H represent polar and hydrophobic residues, respectively (12, 13; see also 14). The association of these motifs with GroEL is difficult to quantitatively analyze due to three levels of heterogeneity (Figure 1.1). First, for every one native SP, there exist multiple misfolded states that form an ensemble of structures, all of which have the potential to interact with GroEL differently. The second level of heterogeneity is topological. Any

one of the misfolded states has several GroEL binding motifs, which can be arranged in multiple ways between the seven SP binding sites of a ring. The third level occurs within the binding pocket itself. Although the seven SP binding pockets of GroEL are all identical, they bind the general sequence P-HHH-P-H which, being redundant, covers a wide range of structural sequences. Thus, within a single SP-GroEL complex, the various SP binding site contacts differ from one another. This heterogeneity makes it difficult to use GroEL's natural SPs to quantitatively characterize their impact on the chaperonin cycle.

We sought to address the SP binding problem by utilizing an SP surrogate with greatly reduced heterogeneity. The stimulatory effects of SP on the chaperonin cycle are rooted in the structural load that bound SP places on the movement of the apical domains within a heptameric ring. Since a typical SP forms multiple (≥2) non-covalent cross-links between the apical domains of GroEL, an effective surrogate should impose similar non-covalent, inter-subunit cross-links. It should also be a single species that interacts with each GroEL subunit in an identical fashion. Since there are seven GroEL subunits, there are numerous topological arrangements for such a probe, and so the number and distribution of the surrogates across the subunits within a ring should be predictable. Finally, the surrogate should create an inter-subunit load that is of uniform energetic value; i.e., the stabilization energy of the inter-subunit crosslink should be identical. These qualities can be found by using non-covalently stacked tetramethylrhodamine dimers as an SP surrogate.

1.1.3 Tetramethylrhodamine Background

Tetramethylrhodamine (TMR) is commonly used in biochemical studies due to its high absorbance and fluorescence in the visible region. It is often covalently attached using a thiol- or amine-reactive linker group. As in previous studies (15-17) we took advantage of the dimer formation or stacking propensity of TMR. Dimer formation is accompanied by a change in the absorbance and fluorescence properties of the dye, a deviation that can be explained by Kasha's exciton coupling

model (18). Fluorescence is forbidden from this state, and large decreases in fluorescence are consistently observed under conditions that favor dimer formation. A blue shift in the peak absorbance indicates that the xanthene planes housing the electronic transition are oriented parallel to one another. Various studies have demonstrated that the xanthenes of the two TMR molecules are stacked in the dimer (16, 19, 20).

TMR has been used as a spectroscopic probe in other biomolecular experiments due to the spectroscopic difference between the dimeric and monomeric dye species. Here we exploit TMR as *both* an allosteric probe and an SP surrogate. The fluorophore can be covalently attached to the protein at a cysteine residue using maleimide chemistry. The self-association of TMR is distance-dependent, and dimers only form when two labeled residues are in close proximity. In order to use TMR dimers as an allosteric probe, the residues to which TMR is attached were chosen such that in the T state they are close enough to allow dimer formation and in the R state, distant enough to prevent dimer formation.

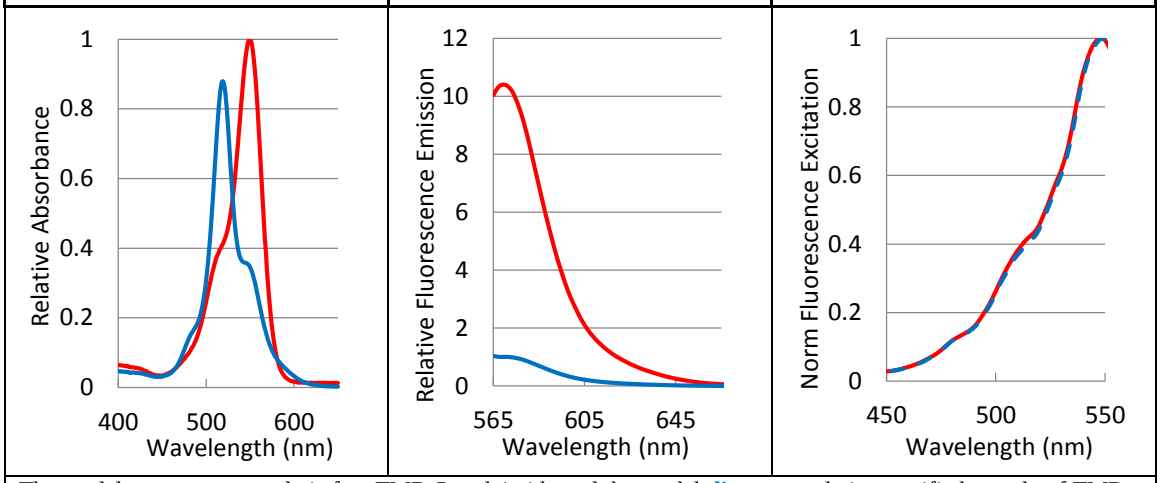
1.2 RESULTS AND DISCUSSION

1.2.1 Tetramethylrhodamine-dithiothreitol as a Model System for Stacked TMR Dimers

The equilibrium of TMR dimer formation (stacking) can be shifted in favor of dimers by lowering the entropic barrier to self-association. To accomplish this TMR-maleimide was reacted with dithiothreitol (DTT) (21) to yield TMR_2-DTT, so as to allow for the concentration-independent formation of dimers. The stacking event is largely driven by favorable van der Waals contacts and the entropy gain associated with the expulsion of water.

The TMR dimer and monomer have distinct absorption and fluorescence spectra (Figure 1.2). TMR alone has an absorbance spectrum that peaks at 551nm with an extinction coefficient of ~75,000 $cm^{-1}\,M^{-1}$. A purified sample of TMR_2-DTT has a comparatively strong peak absorbance at 518nm

Figure 1.2 – Spectroscopic Comparison of TMR Dimers and Monomers



The model **monomer** sample is free TMR-5-maleimide and the model **dimer** sample is a purified sample of TMR_2-DTT. All measurements are taken at room temperature in 50mM Tris-HCl pH 7.5. In plots (A) and (B), the displayed values are relative to the total concentration of dye. (A) The relative molar absorptivity. **Monomer** peaks at 551nm and the **dimer** at 518nm (B) Relative steady state fluorescence emission. The **monomer** is ~ 11X more fluorescent than the model **dimer** sample. (C) The normalized steady state excitation spectra of the two samples. **Monomer** curve is dotted for clarity. Plot (C) demonstrates that the **dimer** is non-fluorescent.

and a shoulder peak around 551nm. The dimer sample is also only weakly fluorescent; an equal concentration of monomer is about 8-13 times more fluorescent.

Although the dominant population in a sample of TMR_2-DTT is the stacked form, it consists of molecules in equilibrium between their stacked and non-stacked conformations. Thus, the fluorescence observed from the TMR_2-DTT sample may either be due to the intrinsic fluorescence of the stacked form, or to the residual population of non-stacked molecules. Evidence suggesting that the stacked form is non-fluorescent can be found by comparing the monomer and dimer fluorescence excitation spectra. The fluorescence excitation spectrum of purified TMR_2-DTT is identical to the absorption spectrum of the monomer (Figure 1.2.C) i.e., the absorbance of the peak at 518nm makes no contribution to the fluorescence of the dimer. The fluorescence of TMR_2-DTT thus reflects the residual population of non-stacked molecules. Since the total concentration of dye molecules in a sample of TMR_2-DTT is known from its absorbance in the presence of SDS, and the size of the non-stacked population is known from the fluorescence, the size of the population

in the stacked form and thus the equilibrium constant for stacking can be calculated. The equilibrium constant allows us to calculate the free energy of formation of the dimer, ΔG_{dim}. The average free energy of stacking is calculated to be -2.8 ± 0.3 kJ/mole dimer (± SEM). Although the calculated free energy is unique to the TMR_2-DTT system, it can be used as a reference for the formation of TMR dimers in the EL242C-TMR system.

1.2.2 Criteria for Positioning the TMR Probe

If an enzyme is labeled with TMR such that in one allosteric state, labeled residues are close enough to form dimers and in another state, too far to associate, the dye can be used as a probe to report the allosteric state of the enzyme. Using simple ball and stick structures of TMR, the maximal C_a - C_a distance between residues that allows for the formation of dimers is ~25Å. In selecting a suitable residue for the placement of TMR, we used the crystal structure of apo GroEL (PDB 1OEL) (22), and the cryo-EM structure of the ATP bound GroEL (PDB 2C7E) (8). These structures were used to find a residue that (i) met the distance criteria for an allosteric probe, (ii) was near the peptide-binding site defined by the cleft between helices H and I, (iii) was solvent accessible and hence amenable to chemical modification and, (iv) whose mutation would not result in enzyme inactivation. Residue K242, which is located at the N terminal end of helix H on the apical domain and is solvent accessible, was selected for mutation and labeling with TMR (Figure 1.3). In the **T** state, the C_α - C_α distance between K242 residues in adjacent subunits is 24.5 Å. Upon ATP binding, this distance increases to 25.4 Å. Mutagenesis was used to create the K242C mutation in the wild type GroEL sequence (23). EL242C can be purified using the same protocol as wild type GroEL. The

Figure 1.3 – GroEL Heptameric Ring

(A) T State	(B) R State

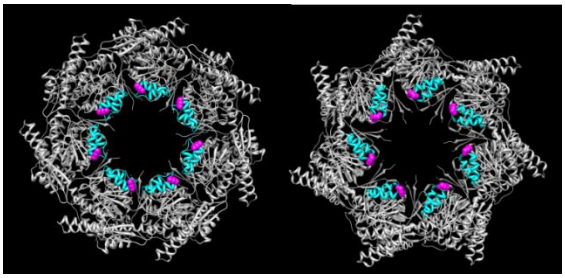


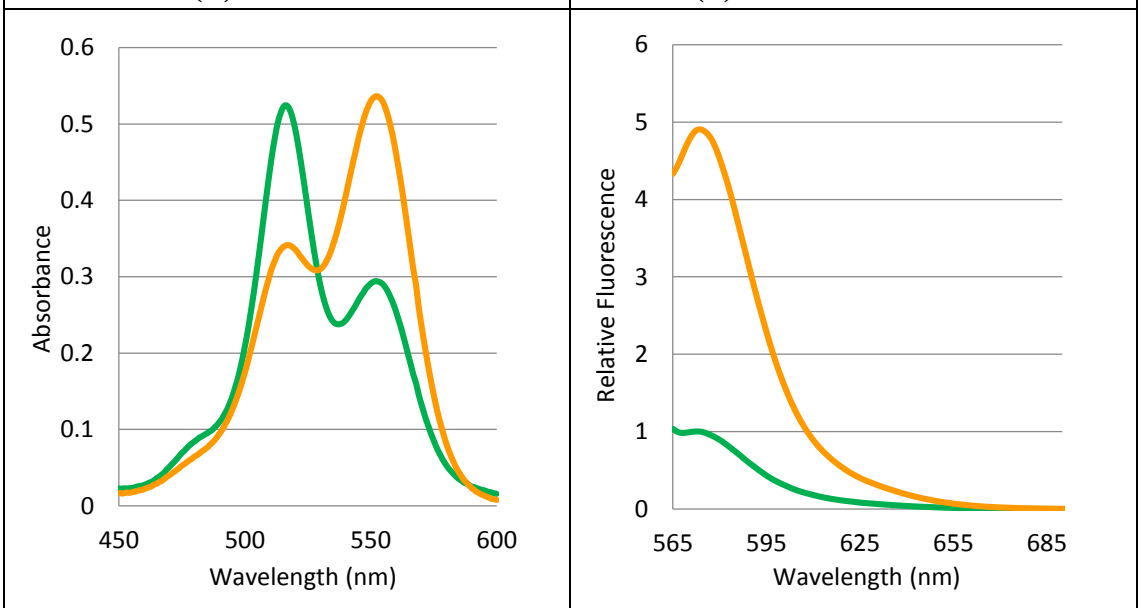
Image of the GroEL heptameric ring looking down the seven-fold axis of symmetry. Highlighted are the **SP binding sites**, located between helices H and I, and **residue 242**, which was subject to mutagenesis. Note the change in orientation of the SP binding sites. T state image from PDB 1OEL and R state image from PDB 2C7E.

location of residue 242 near the SP binding pocket allowed formation of stacked TMR dimers between adjacent subunits in the T allosteric state and their disruption upon addition of ATP. Moreover, as homogeneous non-covalent, inter-subunit cross-links, the TMR dimers serve as SP surrogates.

1.2.3 TMR as an Allosteric Reporter

TMR has been employed to report the allosteric state of GroEL. Figure 1.4 shows the absorption and fluorescence spectra of a sample of EL242C-TMR with an average of 4.3 dyes per ring. In the absence of ATP, GroEL adopts the T state, and the presence of a strong absorption peak at 518nm indicates the presence of a population of TMR dimers. In the presence of ATP, the allosteric equilibrium is shifted in favor of the R state. As supported by both absorbance and fluorescence measurements, the allosteric transition increases the C_α - C_α distance between 242C residues in adjacent subunits, leading to the dissociation of the TMR dimers into monomers. In the presence

Figure 1.4 – TMR Dimers as Allosteric Probes



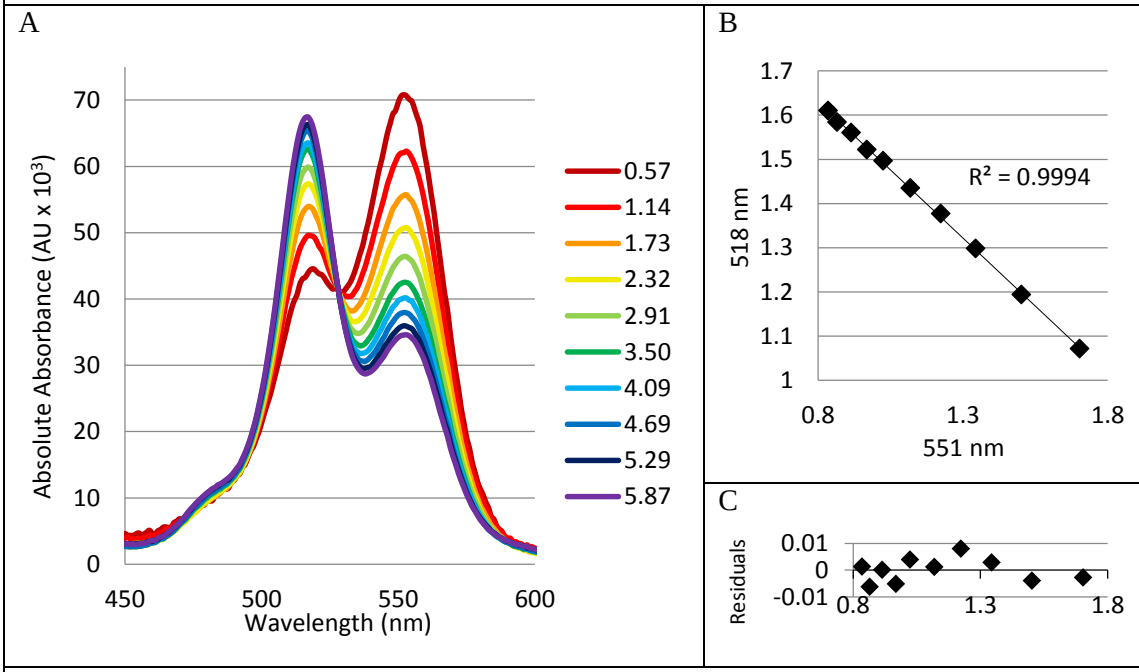
The absorbance (A) and steady state fluorescence emission (B) of a sample of EL242C-TMR with 4.3 TMR per ring is examined in the **absence** and **presence** of 500µM ATP. In the **absence of ATP**, the enzyme is in the T state, indicated by the population of dimers. **Addition of ATP** induces the transition to the R state, and the dimers dissociate into monomers, causing a change in the peak absorbance and increase in the fluorescence.

of ATP, the 518nm dimer peak drops and the 551nm monomer peak rises. Likewise, addition of ATP is marked by a significant increase in fluorescence. Thus, qualitatively, TMR dimers are a suitable allosteric reporter.

In order to use TMR dimers as quantifiable allosteric probes, we determined the size of the population of the individual dye species in a given sample. To accomplish this, we first demonstrated that, in a sample of EL242C-TMR, there exists only two spectrally distinct species. The two species model implies that, in every sample, the entire population of dye molecules can be classified as either dimers or monomers. Mathematically, this can be stated as the sum of the mole fraction of monomer and dimer, α_M and α_D, respectively, is equal to unity, Eq 1.1.

$$\alpha_M + \alpha_D = 1 \qquad \text{Eq 1.1}$$

Figure 1.5 – Two Species Model and Extent of Labeling



TMR bound to GroEL can be classified into two spectrally distinct species implying that bound dye molecules can only exist as either monomers or dimers; no higher order aggregates exist. (A) Relative absorbance of various samples of EL242C-TMR containing the posted number of dye molecules per ring. (B) A plot of the absorbance at 551nm vs 518nm for each curve in (A). (C) The residuals calculated from fitting the plot with a line. The straight line indicates that the change in the size of the population of one species is equally and oppositely accounted for by a change in the size of the other.

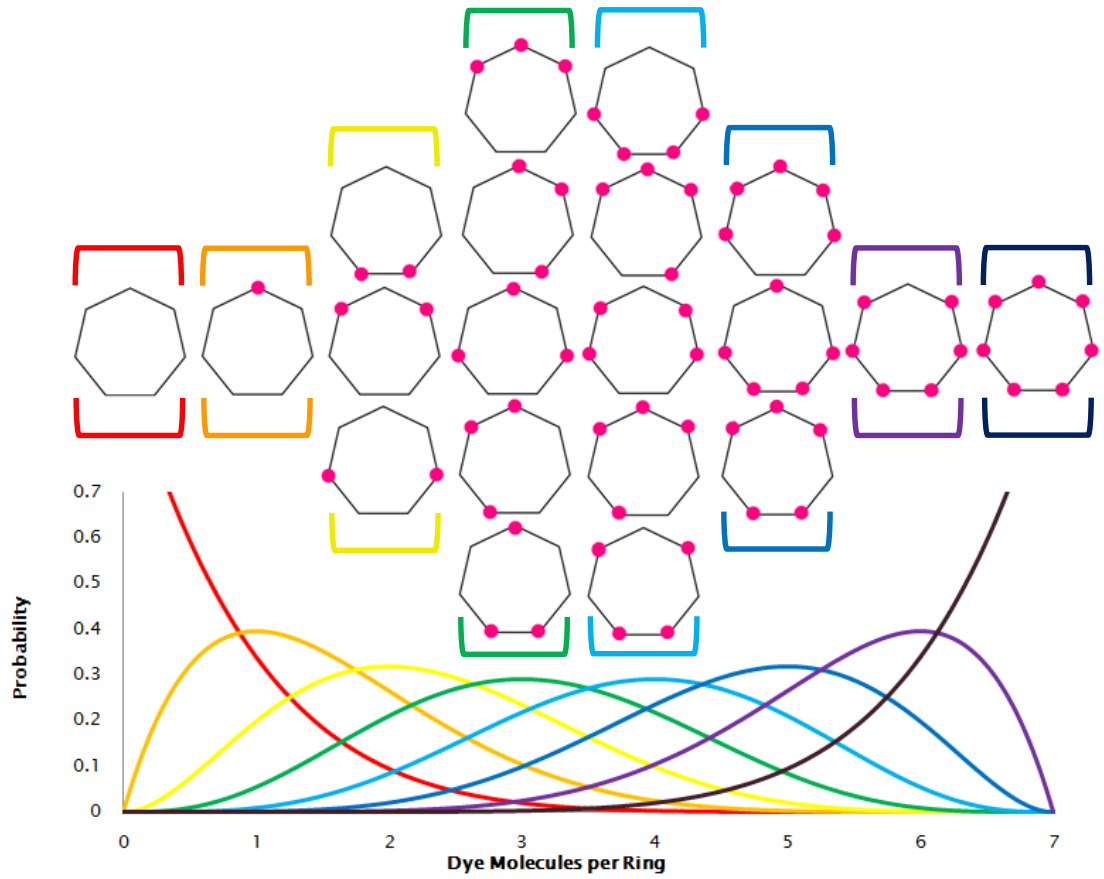
A previous study has indicated that, in solution, concentration-dependent self-association events proceed beyond dimer formation and higher order aggregates are formed (24). The presence of higher order aggregates in the EL242C-TMR system would preclude a simple two species model. The validity of the two-species model was tested by analyzing a series of relative absorption spectra (Figure 1.5.A). If the two species model is correct, then relative changes in the population of one species will be compensated by an equal and opposite change in the other species. Plotting the relative absorbance of each sample at 518nm versus 551nm demonstrates a linear relationship (Figure 1.5.B and 1.5.C). Principal component analysis using singular value decomposition was also used to validate the two-species model (Appendix A.1.6). Thus, we conclude the total population of dye molecules on EL242C can be classified as either monomers or dimers and there are no higher order aggregates.

It is also important to ensure that the spectral properties of the population of monomers in the T state are equivalent to the population of dissociated dimers in the R state. The addition of ATP to a sample of EL242C-TMR shifts the allosteric equilibrium in favor of the R state, causing the dissociation of dimers and increasing the population of monomers. As time elapses, the ATP is hydrolyzed to ADP. This slowly shifts the allosteric equilibrium in favor of the T state, increasing the population of dimers. Figure A.1.1 shows the results of this experiment, and the validity of the two species model. A plot of the absorbance at 551nm verse 518nm shows a strong linear relationship (R^2=0.9998). For the purposes of subsequent analysis we will assume the spectral properties of the monomer in the T and R states are equivalent. Principal component analysis using singular value decomposition also lends credibility to the simple two species model (Appendix A.1.6). Thus, in the T state, dye molecules bound in close proximity spontaneously associate, forming weakly fluorescent dimers, which absorb maximally at 518nm. The addition of ATP triggers the T to R allosteric transition, causing dissociation of the dimers into monomers, marked by a fluorescence increase and exchange between the 518nm and 551nm absorbance peaks.

1.2.4 Stochastic Binding Model

Each EL242C heptameric ring contains 7 possible TMR binding sites, such that a sub-stochiometric addition of TMR creates an ensemble of different structures. In fact, there are 20 possible distinct orientations of dye molecules on a ring (Figure 1.6, *top*). There is evidence that only cysteines at 242 are labeled (Appendix A.1.5). We assume that the labeling process is

Figure 1.6 – The Stochastic Binding Model



The stochastic binding model allows us to quantify the distribution of labeled states. (Top) An illustration of the 20 distinct labeled states in the system. (Bottom) The probability of selecting any one labeled state from the total population can be obtained from the binomial distribution function (Eq. 1.1 and 1.2). Each curve in the binomial distribution corresponds to the proportion of the total population of species within the matching column of labeled states. Individual labeled states within a column are all equally probable.

stochastic: that all binding events are independent and occur with equal probability. This permits the development of the stochastic binding model (Figure 1.6, *bottom*) that describes the distribution of labeling states in any sample from the measured average number of dye molecules per ring, ρ.

The distribution of states is described by a binomial distribution (Eq 1.2 and 1.3). Given the average number of dye molecules per ring, the stochastic binding model gives the fraction of the population containing *k* dye molecules per ring, where *k* = 0-7 in integer values. Each of the labeled states corresponding to *k* dye molecules per ring are equally likely.

$$b_k(\rho) = \frac{7!}{(7-k)!\,k!}\left(\frac{\rho}{7}\right)^k\left(1-\frac{\rho}{7}\right)^{7-k} \qquad \text{Eq 1.2}$$

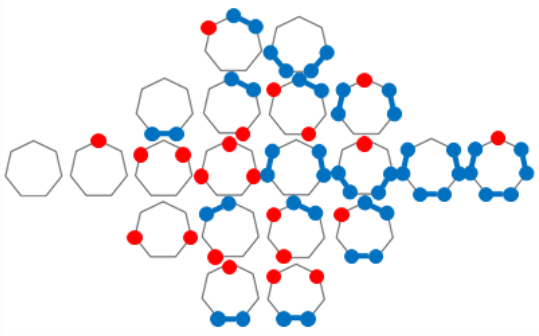
$$\vec{B}(\rho) \equiv \begin{bmatrix} b_0(\rho) \\ \vdots \\ b_7(\rho) \end{bmatrix} \qquad \text{Eq 1.3}$$

1.2.5 Exclusive Nearest Neighbor Assumption

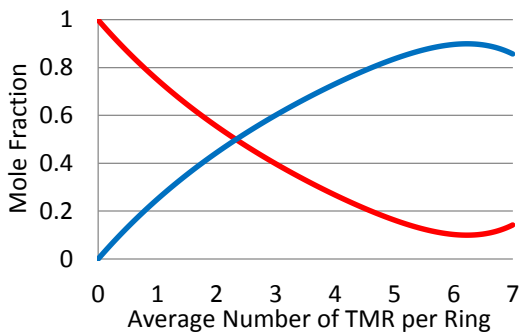
The stochastic binding model allows us to convert a single measured value, the average number of dye molecules per ring, to the fraction of the population in each of the 20 distinct labeled states. We further assume that only adjacent subunits are able to form dimers and that, at any given moment, a single TMR can only be associated with one other TMR: the exclusive nearest neighbor assumption. This assumption is justified considering the geometric constraints imposed by the relatively rigid GroEL structure, and the distance constraint of dimer formation. The exclusivity in the assumption is supported by the observation that no higher order aggregates exist in the sample.

Figure 1.7 – Combining the Stochastic Binding Model and the Exclusive Nearest Neighbor Assumption

A



B



Each of the 20 distinct labeled states can be assigned a fractional contribution to the population of **monomers** and **dimers**. These weights are collected into the matrix $\widehat{M_F}$. (B) The matrix product $\widehat{M_F} \cdot \vec{B}(\rho)$ yields two equations, the mole fraction of **monomers** and the mole fraction of **dimers**, as a function of TMR/ring, ρ.

The nearest neighbor assumption allows us to attach, to each of the 20 distinct labeled states, its fractional contribution to the monomer and dimer populations (Figure 1.7.A).

Mathematically, the fractional contribution of the two species is accounted for in the matrix $\widehat{\mathrm{M}_F}$. The first and second rows in $\widehat{\mathrm{M}_F}$ corresponds to the fractional contribution to the monomer and dimer population, respectively. Taking the two assumptions together, one can predict the mole fraction of the population in the dimer, $\alpha_D(\rho)$, and monomer form, $\alpha_M(\rho)$ given an average number of dye molecules per ring, ρ (Eq 1.4, Figure 1.7.B).

$$\widehat{\mathrm{M}_F} \equiv \begin{bmatrix} 0 & 1 & {}^{2}/_{3} & {}^{7}/_{15} & {}^{6}/_{20} & {}^{1}/_{5} & 0 & {}^{1}/_{7} \\ 0 & 0 & {}^{1}/_{3} & {}^{8}/_{15} & {}^{14}/_{20} & {}^{4}/_{5} & 1 & {}^{6}/_{7} \end{bmatrix}$$

$$\widehat{\mathrm{M}_F} \cdot \vec{B}(\rho) = \begin{bmatrix} \alpha_M(\rho) \\ \alpha_D(\rho) \end{bmatrix} \qquad \text{Eq 1.4}$$

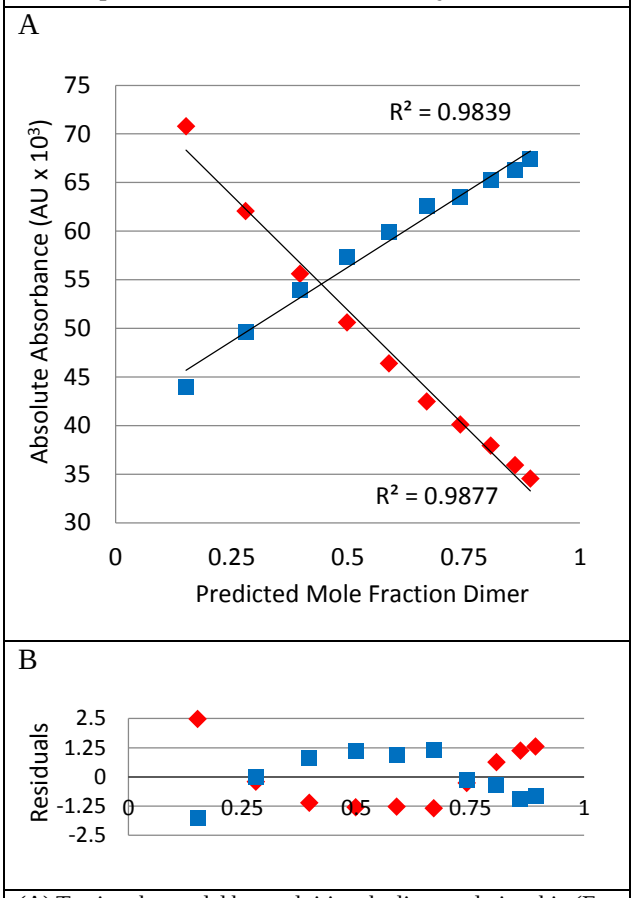
The proposed model predicts the change in the mole fraction of monomers and dimers with the extent of labeling (Eq 1.4). To test this model we exploited the linear relationship between the mole fraction of dimer and the relative absorbance of the sample, found by combining the two species model (Eq 1.1) with Beer's Law for two species (Eq 1.5) to yield Eq 1.6.

$$A_{rel} = \epsilon_M \, \alpha_M(\rho) + \epsilon_D \, \alpha_D(\rho) \qquad \text{Eq 1.5}$$

$$A_{rel} = (\epsilon_D - \epsilon_M) \, \alpha_D(\rho) + \epsilon_M \qquad \text{Eq 1.6}$$

The plot of the calculated mole fraction of dimer against the observed absorbance at 518nm and 551nm was linear (Figure 1.8.A). However, the non-random nature of the residuals points to a small source of error of unknown origin (Figure 1.8.B). This can be explained by invoking positive cooperativity in the labeling process; i.e., a dye molecule has a slight preference to react with a vacant subunit that is adjacent to an already labeled subunit as opposed to reacting in a strictly stochastic manner.

Figure 1.8 – Testing the Stochastic Binding Model and the Exclusive Nearest Neighbor Assumption in the EL242C-TMR System



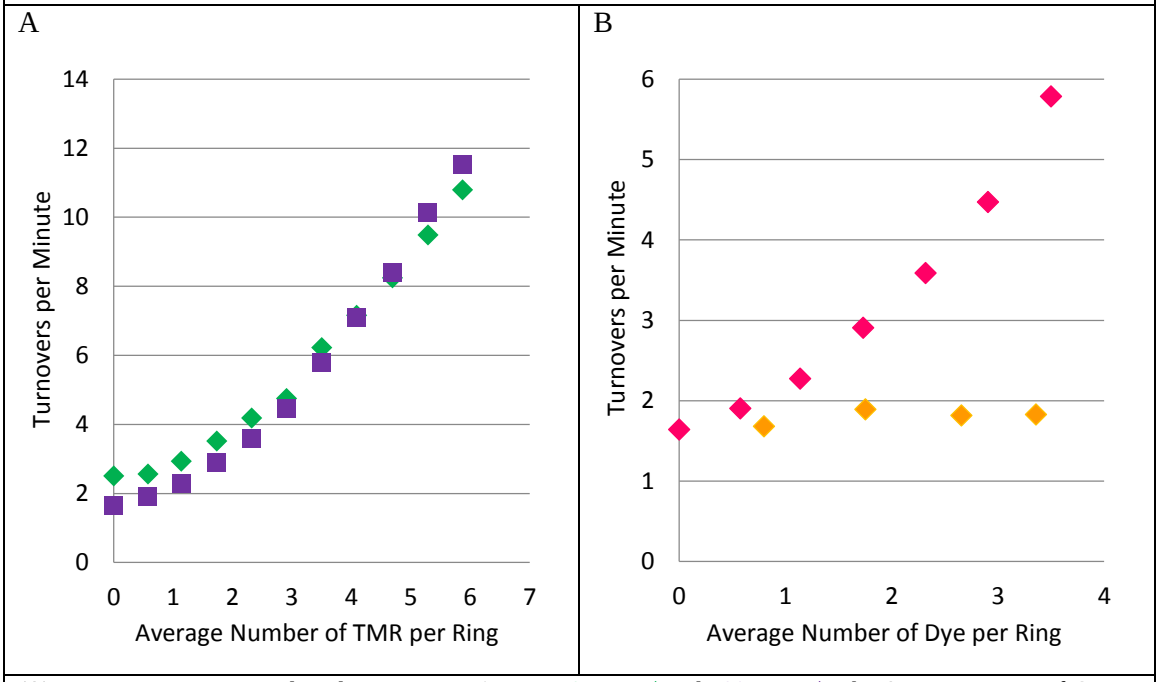
(A) Testing the model by exploiting the linear relationship (Eq. 1.6) between the predicted mole fraction of **dimer** (and **monomer**), $\alpha_D(\rho)$, and the relative absorbance (Figure 1.5) of a samples of EL242C-TMR at **551nm** and **517nm**. (B) The fit is relatively good agreement, R^2=0.98, but the residuals are not randomly distributed.

1.2.6 TMR as an SP Surrogate

SP binds to the T state of GroEL and stimulates its ATPase by placing a load on the apical domains of the ring. The bound SP serves as a non-covalent tether between two adjacent subunits. This stabilizes the T state, which has weak affinity for nucleotide, accelerating the rate of product release, and increasing the rate of ATP consumption (10, 11). Likewise, the apical domains of the EL242C-TMR are tethered by the non-covalent stacking of the dye molecules. And similarly, an increase in the steady state turnover rate is observed with increasing number of dimers per ring (Figure 1.9.A). This supports the assertion that TMR dimers behave as surrogates for SP in the chaperonin cycle. The stimulation is due to the cross-linking effect of TMR dimers, and not merely to the presence of a hydrophobic fluorophore close the SP binding site. To address this EL242C was labeled with fluoresceine-5-maleimide (F5M), a compound structurally similar to TMR. However, it neither forms dimers nor stimulates ATPase activity (Figure 1.9.B). The TMR dimers are thus behaving as molecular latches, perturbing the allosteric equilibrium to favor the T state.

In addition to being an allosteric probe, TMR dimers act as an SP surrogate because they serve as non-covalent crosslinks, between subunits within a ring, that place a load on the apical domains in

Figure 1.9 – ATPase Activity of Dye Labeled EL242C in 1mM ATP



(A) Turnovers per minute plotted against TMR/ring at **10mM K^+** and **100mM K^+**. The ATPase activity of GroEL increases as the mole fraction of dimers increases. (B) Turnovers per minute plotted against dye/ring for EL242C labeled with **TMR** and **F5M**, in the presence of 100mM K^+. The increase in the turnover rate is not due to the mere presence of fluorophore, as indicated by the lack of ATPase stimulation observed in samples of EL242C-**F5M**.

the same way as SP. TMR serves as an ideal SP surrogate because it circumvents the problems of heterogeneity inherent in the SP binding problem (Figure 1.1). As opposed to SP, TMR is a small molecule that covalently binds to a single position near the SP binding site. In demonstrating the use of TMR as an allosteric probe, we showed that the topological arrangement is readily predictable using the stochastic binding model. The proportion of the bound TMR that is in the dimeric state can be computed with the exclusive nearest neighbor assumption. Finally, each non-covalent dimeric stacking bond is of identical strength. Since the T to R transition is concerted among the subunits within a ring, the energetic contribution to the stabilization of the T state is a simple sum of the number of dimers per ring. A ring can contain from zero to three dimers, and thus by increasing the number of dimers per ring, we can vary the cross-linking stabilization energy that shifts the allosteric equilibrium.

1.2.7 An Expanded Nested Cooperativity Model (NCM) of GroEL

The negative and positive cooperativity displayed by GroEL during nucleotide binding requires a hybrid MWC/KNF type allosteric model known as the Nested Cooperativity Model. As originally formulated (25, 26) the NCM contained a simplifying assumption that ATP bound exclusively to the R state. Subsequent work, however, showed that the T state could also bind and hydrolyze ATP. (10, 26) An expanded version of the NCM, without the exclusive binding assumption was developed by Gresham (27) and was used to quantitatively analyze the turnover rate of GroEL (Eq 7). The original formulation can be recovered from the expanded form by setting the *c* equal to zero. It contains seven independent parameters that are to be fit for a given set of conditions. At its core, the NCM is an allosteric partition function of GroEL that accounts for the positive intra-ring cooperativity and negative inter-ring cooperativity. The equation gives the fraction of the population with ATP bound, in each of the three allosteric states, as a function of ATP concentration. In its expanded form, this fraction is multiplied by the allosteric state's V_{max}, and the sum of these three terms gives the steady state turnover rate. Two other parameters include the dissociation constants, k_T and k_R, of ATP from a T ring and an R ring, respectively. The two terms appear in the NCM as $c = {k_R}/{k_T}$ and $\alpha = {[ATP]}/{k_R}$. Lastly, the NCM contains two allosteric equilibrium terms, L_1 for the TT to TR transition, and L_2 for the TR to RR transition.

$$v = \frac{V_{TT}\, c\, \alpha\, (1 + c\, \alpha)^{13} + \frac{1}{2} V_{TR}\, L_1 [c\, \alpha\, (1 + c\, \alpha)^6 (1 + \alpha)^7 + \alpha\, (1 + c\, \alpha)^7 (1 + \alpha)^6] + V_{RR}\, L_1\, L_2\, \alpha\, (1 + \alpha)^{13}}{(1 + c\, \alpha)^{14} + L_1\, (1 + c\, \alpha)^7 (1 + \alpha)^7 + L_1 L_2 (1 + \alpha)^{14}} \qquad \text{Eq 1.7}$$

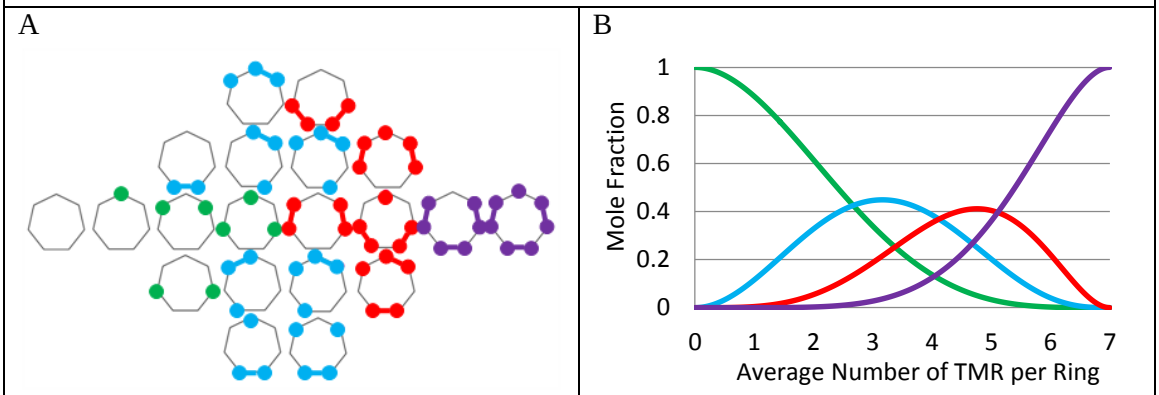
In our analysis, the NCM (Eq 1.7) is modified to predict the distribution of allosteric states as a function of ATP concentration. Doing so decreases the number of independent parameters, replacing the three V_{max} terms in the expanded derivation, with a single proportionality constant, A. This constant converts the instrument's arbitrary fluorescence readout value to the fluorescence increase observed from the dissociation of a dimer, upon transition of a ring to the R

state. Given a single GroEL 14mer with *m* TMR dimers on one ring, and *n* TMR dimers on the other, there will be an increase proportional to *m* upon the TT to TR and a fluorescence increase proportional to *n* upon the TR to RR transition. The modified NCM, Eq 8, predicts the observed fluorescence increase in a labeled species with *m* dimers on the first ring, and *n* on the second.

$$F' = \frac{m\,A\,L_1\,\frac{1}{2}\,[c\,\alpha\,(1+c\,\alpha)^6(1+\alpha)^7 + \alpha\,(1+c\,\alpha)^7(1+\alpha)^6] + (m+n)A\,L_1\,L_2\,\alpha\,(1+\alpha)^{13}}{(1+c\,\alpha)^{14} + L_1\,(1+c\,\alpha)^7(1+\alpha)^7 + L_1\,L_2\,(1+\alpha)^{14}} \qquad \text{Eq 1.8}$$

As pointed out above, each sample consists of an ensemble of labeled states. The stochastic binding model and the exclusive nearest neighbor assumption were used to calculate the size of each labeled state, in a given sample, as a function of the average number of dye molecules per ring. Each state can be binned corresponding to the number of dimers it contains (Figure 1.10.A). Mathematically, this function is performed by the matrix $\hat{\mathbb{C}}$. The matrix product of $\hat{\mathbb{C}}$ and $\vec{B}(\rho)$ (Eq 1.3) produces 4 functions, $c_m(\rho)$, where m is an integer between zero and three (Figure 1.10.B, Eq 1.9). These functions give the size of the population of the four cross-linked states (m = 0-3 per ring) as a function of the average number of dye molecules per ring.

Figure 1.10 – Distribution of Cross-Linked States



(A) Illustration of how the 20 distinct labeled states can be assigned to one of the four cross-linked states: **0 dimers/ring**, **1 dimer/ring**, **2 dimers/ring** and, **3 dimers/ring.** These weights are collected in the matrix $\hat{\mathbb{C}}$. (B) The matrix product $\hat{\mathbb{C}} \cdot \vec{B}(\rho)$ yield four equations, $c_m(\rho)$ where m is an integer between zero and three, that correspond to the size of the population of heptameric rings with m dimers.

$$\hat{\mathrm{C}} \equiv \begin{bmatrix} 1 & 1 & {}^{2}/_{3} & {}^{1}/_{5} & 0 & 0 & 0 & 0 \\ 0 & 0 & {}^{1}/_{3} & {}^{4}/_{5} & {}^{3}/_{5} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & {}^{2}/_{5} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\hat{\mathrm{C}} \cdot \vec{B}(x) = \begin{bmatrix} c_0(x) \\ c_1(x) \\ c_2(x) \\ c_3(x) \end{bmatrix} \qquad \text{Eq 1.9}$$

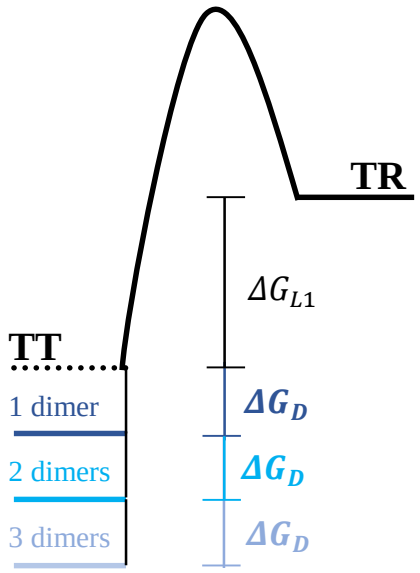
It is known that SP places a load on the apical domains of a ring (30), and that the presence of SP leads to a stimulation in ATPase activity (11). In order to arrive at a quantitative impact of SP on the chaperonin cycle, we view the load that SP places on the ring, to be stabilization of the T state and thus an increase in the free energy of the T to R allosteric transition (Figure 1.11.A). Likewise, the cross-linking effect of the stacking interaction in TMR dimers stabilizes the T state of GroEL. The advantage of the TMR dimer is that it is of uniform energetic value and likely to be of a comparable magnitude to the TMR_2-DTT stacking free energy: -2.7 ± 0.3 kJ/mole dimer. Furthermore, a ring can have from zero to three TMR dimers. Because of the concerted nature of the allosteric transitions exhibited by GroEL, the energetic contribution of the stacked dye molecules to the stabilization of the T state can be assumed to be additive across a ring. The stabilization of an allosteric state quantitatively enters into the NCM as a modification to the allosteric equilibrium terms, L_1 and L_2 (Figure 1.11.B). The end result is a modified form of the NCM that predicts the biphasic fluorescence increase observed for a given labeled state, taking into account the perturbation placed on the allosteric equilibrium by the presence of the dimers (Eq 1.10).

$$F_{mn} = \frac{m\,A\,L_1\,\mathcal{D}^m \frac{1}{2}\,[c\,\alpha\,(1+c\,\alpha)^6(1+\alpha)^7 + \alpha\,(1+c\,\alpha)^7(1+\alpha)^6] + (m+n)\,A\,L_1\,L_2\,\mathcal{D}^{m+n}\,\alpha\,(1+\alpha)^{13}}{(1+c\,\alpha)^{14} + L_1\,\mathcal{D}^m\,(1+c\,\alpha)^7(1+\alpha)^7 + L_1\,L_2\,\mathcal{D}^{m+n}(1+\alpha)^{14}} \qquad \text{Eq 1.10}$$

$$\langle F \rangle = \sum_{n=0}^{3} \sum_{m=0}^{3} C_m\,C_n\,F_{mn} \qquad \text{Eq 1.11}$$

Figure 1.11 – Quantifying the Effect of SP on the Allosteric Equilibrium

(A) Free Energy Diagram of Allosteric Transition



$$\Delta G = \Delta G_{L1} + n\Delta G_D$$

(B) Derivation of NCM Modification

$$K_{TT \to TR} = e^{-\frac{\Delta G}{RT}}$$

$$= e^{-\frac{\Delta G_{L1} + n\Delta G_D}{RT}}$$

$$= e^{-\frac{\Delta G_{L1}}{RT}} e^{-\frac{n\Delta G_D}{RT}}$$

$$= L_1 * \left(e^{-\frac{\Delta G_D}{RT}}\right)^n$$

$$= L_1 * \mathcal{D}^n$$

Where

$$\mathcal{D} \equiv \left(e^{-\frac{\Delta G_D}{RT}}\right)$$

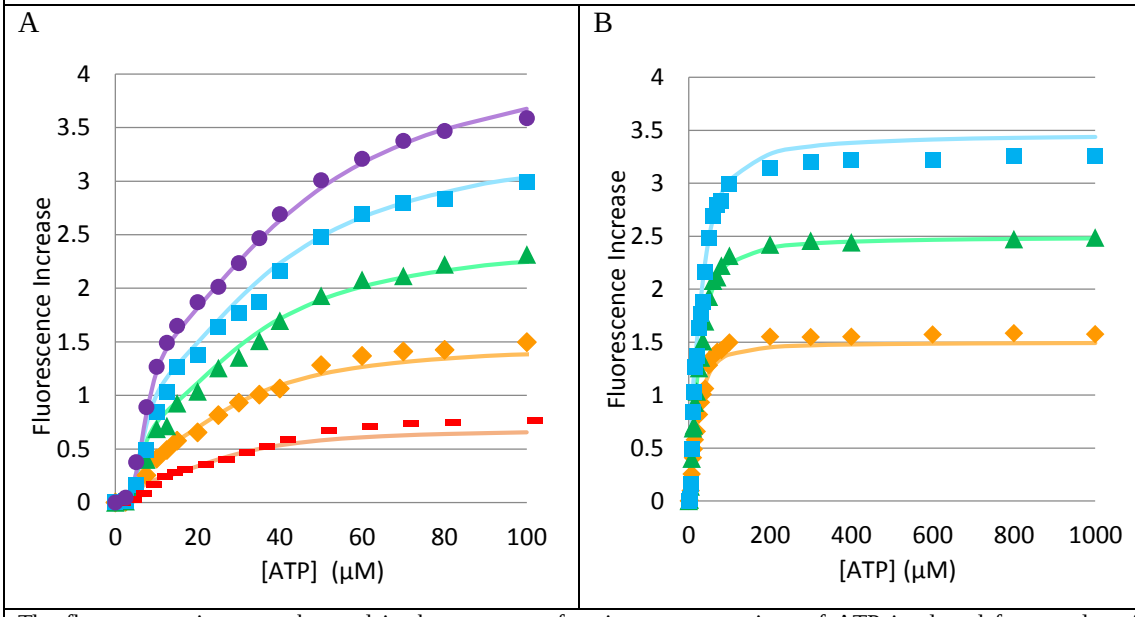
And

$$\Delta G_D = -\Delta G_{Dim}$$

(A) A free energy diagram of GroEL's first allosteric transition, TT to TR. The diagram indicates how the presence of a dimer on a ring stabilizes the T state and increases the free energy gap between the TT and TR state. Since the allosteric transition is concerted, the effect of multiple dimers is additive. (B) Mathematically, this interpretation leads us to quantitatively account for the presence of TMR as a perturbation to the allosteric equilibrium.

Combining the modified NCM in Eq 1.10 with the statistical distribution of cross-linked states (Eq 1.9) yields an equation that predicts the fluorescence increase associated with the allosteric transitions, in a sample with a known average number of dye molecules per ring, as a function of ATP concentration (Eq 1.11). The solution to Eq 1.11 depends on the value of the various independent parameters in the equation, *A*, *c*, k_r, L_1, L_2 and, *D*. The large number of parameters gives the equation a great deal of flexibility when being fit to a particular data set. However, we have collected data sets for multiple samples, with varying numbers of average dye molecules per ring. These data sets were fit globally with Eq 1.11 (Figure 1.12). Varying any one parameter will change the fit for all data sets, allowing for increased confidence in the fitted values.

Figure 1.12 – Fluorescence Increase versus [ATP]



The fluorescence increase observed in the presence of various concentrations of ATP is plotted for samples of EL242C-TMR with an average of **1.8 TMR/ring**, **2.8 TMR/ring**, **3.8 TMR/ring**, **4.7 TMR/ring** and, **5.6 TMR/ring**. (A) [ATP] range from 0-100μM. (B) [ATP] range from 0-1mM. Two of the data sets, **1.8 TMR/ring** and **5.6 TMR/ring**, were collected over an [ATP] range of 0-100μM and the remaining three sets over an [ATP] range of 0 - 1mM. Solid lines are the results of the global fit of the data sets to the expanded NCM (Eq. 1.11).

Table 1.1 – Results of the Global Fit to the Modified NCM

Parameter	Value
A	0.98 ± 0.10
kr	0.9 ± 0.3 μM
c	0.042 ± 0.008
ΔG_{L1}	26 ± 11 kJ/mol
ΔG_{L2}	46 ± 2 kJ/mol
ΔG_{D}	2.6 ± 1.0 kJ/mol

Results of the global fit of the expanded NCM (Eq. 11) to the fluorescence verse [ATP] data. See text for explanation of fitted parameters. The range of each parameter that falls within the 95% confidence interval of the global fit is given.

Table 1.2 – Evaluating Goodness of Global Fit

TMR/Ring	R^2 Value
1.79	.9115
2.78	.9776
3.78	.9909
4.68	.9828
5.49	.9966

The R^2 value of each curve (Figure 1.12) in the global fit of fluorescence verse [ATP] data to the expanded NCM (Eq. 11).

The results of the global fit to Eq 1.11 are summarized in Table 1.1. The indicated range of each parameter falls within the 95% confidence interval of the fit. The free energy of the TT to TR and

TR to RR transitions were found to be 27 ± 11 and 46 ± 2 kJ/mol, respectively. TMR dimers acting as an SP surrogate increase this free energy gap by 2.6 ± 1.0 kJ/mol dimer, a value that is similar to that for the dissociation of TMR_2-DTT. These experiments demonstrate that the GroEL nano-machine is able to perform work and overcome a load of at least 7.8 kJ/mole ring. This supports the idea of GroEL function that relies on forced unfolding of bound substrate protein, in order to free it from a kinetic folding trap, affording it another opportunity to reach its native state.

1.3 CONCLUDING REMARKS

1.3.1 TMR as an Allosteric Reporter for GroEL

TMR, covalently bound near the SP binding site of GroEL's subunits, forms populations of monomers and dimers depending on nearest neighbor occupancy. Combining the stochastic binding model and the exclusive nearest neighbor assumptions, allows us to predict the distribution of labeled states from the average number of TMR molecules per ring. Introducing ATP induces an allosteric conformational change that causes dimers to dissociate into monomers. Since the monomers and dimers have different spectroscopic signatures, the transition can be monitored using either absorbance or fluorescence spectroscopy. The change in the observed fluorescence or absorbance, upon the addition of ATP, can be directly correlated with a change in the population of the T and R allosteric states, providing a quantitative measure of the allosteric activity.

1.3.2 TMR Dimers as Surrogates for SP

In the T state, TMR labeled 242C residues are close enough to self-associate into dimers. These dimers act as non-covalent intra-ring, inter-subunit crosslinks that place a load on the apical domain, mimicking the effects of the crosslinking force provided by SP. Consequently, SP and the TMR dimers stimulate the ATPase activity of GroEL by stabilizing the T state with this crosslinking force. However, TMR dimers avoid the problems associated with the heterogeneous

nature of SP binding to GroEL. The predictable distribution of dimers and the uniformity of the crosslinking force, make TMR dimers a suitable surrogate for quantifying the role of SP in the catalytic cycle of GroEL.

The crosslinking force of the dimer increases the free energy gap in the T to R allosteric transition by an amount equal to the free energy of dimer formation. The effect of this stabilization can be factored into the NCM as a modification to the allosteric equilibrium terms, L_1 and L_2. The NCM can also be modified for use as an allosteric partition function, to predict fluorescence increases that correlate with allosteric population changes instead of turnover rates. When the modified NCM is paired with the predicted distribution of labeled states the parameters in the equation can be globally fit to a series of data sets allowing for increased confidence in the fitted parameters.

The proposed model highlights the importance of the SP cross-linking energy within the allosteric cycling of GroEL. The load placed on the apical domain stabilizes the T state, increasing the free energy gap and shifting the allosteric equilibrium in favor of the T state. Because of negative cooperativity, this effect is more significant for the TR to RR transition than the TT to TR transition. The model also supports the iterative annealing model of GroEL function. GroEL performed work on the TMR dimers, overcoming the free energy of dimer formation while transitioning from the T to R state. Likewise, the iterative annealing model suggests that GroEL can perform work on SPs by forcibly unfolding the bound SP upon its release into the central cavity, rescuing it from its kinetic folding trap and affording it another opportunity to reach its native state.

1.4 MATERIALS AND METHODS

1.4.1 Experimental Materials

1.4.1.1 Instrumentation

See Appendix A.1.

1.4.1.2 Construction of GroEL K242C

The mutant K242C was prepared as described in (23) and in supporting information (Appendix A.1.2) Materials and Methods. Note that we have inserted the K242C mutation into a wild-type background rather than into a cysteine-free background because, in our hands, we have observed cysteine-free GroEL is much less stable than the wild type.

1.4.1.3 Purification of GroEL

Wild-type GroEL, the mutant GroEL K242C, free of tryptophan containing contaminants, were purified as previously described (11). Typically, GroEL preparations contained <0.2 moles of contaminating substrate protein (SP)/ GroEL14; i.e. <10% of the rings may be contaminated with an ensemble of SPs. The concentration of purified EL242C was calculated by measuring the absorbance at 280nm and using the extinction coefficient of 9,600$cm^{-1}M^{-1}$.

1.4.1.4 Synthesis and Purification of TMR_2-DTT

See Appendix A.1.

1.4.1.5 Labeling GroEL K242C

See Appendix A.1.

1.4.1.6 Steady-state ATP Hydrolysis

The steady-state hydrolysis of ATP was measured at 37°C as previously described (10).

1.4.1.7 Fluorescence verse ATP

See Appendix A.1.

1.4.2 Computational Materials

1.4.2.1 Data Analysis

Data from all instruments was saved in ASCII format and imported into Microsoft Excel 2010. All data analysis, except singular value decomposition (Appendix A.1.6) was performed in Excel. Globally fitting the fluorescence verse ATP experimental data to Eq 1.11 was performed by non-linear least squares analysis. The least squares value for each curve was summed and minimized using Excel solver. The 95% confidence interval of the fit was taken to be plus or minus 1.65

standard deviations from the mean of the residuals. Error in fitted parameters was calculated as the largest deviation from the fitted least-squares value that falls within the 95% confidence interval. R^2 values for each curve were calculated as detailed in DeVore (28). SVD, performed in MATLAB, was used in a principle component analysis of the spectra of EL242C-TMR in support of the two species model. Over 98% of the variance was accounted by this model (29).

1.5 REFERENCES

1. Thirumalai, D. and G. H. Lorimer. "Chaperonin-Mediated Protein Folding." *Annu Rev Biophys Biomol Struct* 30, (2001): 245-69.

2. Horwich, A. L. and W. A. Fenton. "Chaperonin-Mediated Protein Folding: Using a Central Cavity to Kinetically Assist Polypeptide Chain Folding." *Q Rev Biophys* 42, no. 2 (2009): 83-116.

3. Braig, K., Z. Otwinowski, R. Hegde, D. C. Boisvert, A. Joachimiak, A. L. Horwich and P. B. Sigler. "The Crystal Structure of the Bacterial Chaperonin Groel at 2.8 A." *Nature* 371, no. 6498 (1994): 578-86.

4. Chen, L. and P. B. Sigler. "The Crystal Structure of a Groel/Peptide Complex: Plasticity as a Basis for Substrate Diversity." *Cell* 99, no. 7 (1999): 757-68.

5. Landry, S. J., J. Zeilstra-Ryalls, O. Fayet, C. Georgopoulos and L. M. Gierasch. "Characterization of a Functionally Important Mobile Domain of Groes." *Nature* 364, no. 6434 (1993): 255-8.

6. Xu, Z., A. L. Horwich and P. B. Sigler. "The Crystal Structure of the Asymmetric Groel-Groes-(Adp)7 Chaperonin Complex." *Nature* 388, no. 6644 (1997): 741-50.

7. Horovitz, A., Y. Fridmann, G. Kafri and O. Yifrach. "Review: Allostery in Chaperonins." *J Struct Biol* 135, no. 2 (2001): 104-14.

8. Ranson, N. A., G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich and H. R. Saibil. "Atp-Bound States of Groel Captured by Cryo-Electron Microscopy." *Cell* 107, no. 7 (2001): 869-79.

9. Viitanen, P. V., T. H. Lubben, J. Reed, P. Goloubinoff, D. P. O'Keefe and G. H. Lorimer. "Chaperonin-Facilitated Refolding of Ribulosebisphosphate Carboxylase and Atp Hydrolysis by Chaperonin 60 (Groel) Are K+ Dependent." *Biochemistry* 29, no. 24 (1990): 5665-71.

10. Grason, J. P., J. S. Gresham, L. Widjaja, S. C. Wehri and G. H. Lorimer. "Setting the Chaperonin Timer: The Effects of K+ and Substrate Protein on Atp Hydrolysis." *Proceedings of the National Academy of Sciences* 105, no. 45 (2008): 17334-17338.

11. Grason, J. P., J. S. Gresham and G. H. Lorimer. "Setting the Chaperonin Timer: A Two-Stroke, Two-Speed, Protein Machine." *Proc Natl Acad Sci U S A* 105, no. 45 (2008): 17339-44.Stan G, et al. (2005) Identifying natural substrates for chaperonins using a sequence-based approach. *Protein Society*, *14*:193-201.

12. Stan, G., B. R. Brooks, G. H. Lorimer and D. Thirumalai. "Identifying Natural Substrates for Chaperonins Using a Sequence-Based Approach." *Protein Sci* 14, no. 1 (2005): 193-201.

13. Stan, G., B. R. Brooks, G. H. Lorimer and D. Thirumalai. "Residues in Substrate Proteins That Interact with Groel in the Capture Process Are Buried in the Native State." *Proc Natl Acad Sci U S A* 103, no. 12 (2006): 4433-8.

14. Azia, A., R. Unger and A. Horovitz. "What Distinguishes Groel Substrates from Other Escherichia Coli Proteins?" *FEBS J* 279, no. 4 (2012): 543-50.

15. Hamman, B. D., A. V. Oleinikov, G. G. Jokhadze, D. E. Bochkariov, R. R. Traut and D. M. Jameson. "Tetramethylrhodamine Dimer Formation as a Spectroscopic Probe of the Conformation of Escherichia Coli Ribosomal Protein L7/L12 Dimers." *J Biol Chem* 271, no. 13 (1996): 7568-73.

16. Blackman, M, J Corrie, J Croney, G Kelly, J Eccleston and D Jameson. "Structural and Biochemical Characterization of Fluorogenic Rhodamine Labeled Malarial Protease Substrate." *Biochemistry* 41, (2002).

17. Okoh, M. P., J. L. Hunter, J. E. Corrie and M. R. Webb. "A Biosensor for Inorganic Phosphate Using a Rhodamine-Labeled Phosphate Binding Protein." *Biochemistry* 45, no. 49 (2006): 14764-71.

18. Kasha, M. "Energy Transfer Mechanisms and the Molecular Exciton Model for Molecular Aggregates." *Radiat Res* 20, (1963): 55-70.

19. Ilich, P., P.K. Mishra, S. Macura and T.P. Burghardt. "Direct Observation of Rhodamine Dimer Structure in Water." *Spectrochimica acta. Part A, Molecular Spectroscopy* 52, no. 10 (1996): 1323-1330.

20. Edmundson, A. B., K. R. Ely and J. N. Herron. "A Search for Site-Filling Ligands in the Mcg Bence-Jones Dimer: Crystal Binding Studies of Fluorescent Compounds." *Mol Immunol* 21, no. 7 (1984): 561-76.

21. Cleland, W. W. "Dithiothreitol, a New Protective Reagent for Sh Groups." *Biochemistry* 3, (1964): 480-2.

22. Braig, K., P. D. Adams and A. T. Brunger. "Conformational Variability in the Refined Structure of the Chaperonin Groel at 2.8 a Resolution." *Nat Struct Biol* 2, no. 12 (1995): 1083-94.

23. Yang, Y. "Site-Directed Mutagenesis of Groel: Developing a System for Monitoring Allosteric Movements by Fluorescence Resonance Energy Transfer (Fret)." (MS Thesis) University of Maryland, 2006.

24. Toptygin, D., B. Packard and L. Brand. "Resolution of Absorption Spectra of Rhodamine 6g Aggregates in Aqueous Solution Using the Law of Mass Action." *Chemical Physics Letters* 5, no. 227 (1997): 430-435.

25. Yifrach, Ofer and A. Horovitz. "Nested Cooperativity in the Atpase Activity of the Oligomeric Chaperonin Groel." *Biochemistry* 34, no. 16 (1995): 5303.

26. Yifrach, O. and A. Horovitz. "Allosteric Control by Atp of Non-Folded Protein Binding to Groel." *J Mol Biol* 255, no. 3 (1996): 356-61.

27. Gresham, J.S. "Allostery and Groel: Exploring the Tenets of Nested Cooperativity." (PhD Thesis) University of Maryland, 2004.

28. DeVore, J. *Probability and Statistics for Engineers and Scientists*. Vol. 7. Belmont: Thomson, 2008.

29. Lay, D. *Linear Algebra and Its Applications*. Vol. 3. Boston: Pearson, 2006.

30. Motojima, Charu Chaudhry, W. A. Fenton, G. W. Farr and A. L. Horwich. "Substrate Polypeptide Presents a Load on the Apical Domains of the Chaperonin Groel." *PNAS* 101, no. 42 (2002): 15005.

Chapter 2

Investigation of the Pre-Steady State Kinetics of Inorganic Phosphate Release in the GroEL/GroES System

2.1 INTRODUCTION

2.1.1 Background

The GroEL/GroES catalytic cycle has received a lot of attention by the scientific community over the past 20 years, due to its crucial role in the survival of the organism, and the complexity of the cycle itself. The deviation of GroEL's activity from the typical hyperbolic dependence on substrate concentration implies the importance of the roles that communications, between subunits within a ring and rings within a complex, play in the cycle (1-4). Thus far, much of the effort towards understanding these catalytic anomalies have focused on the first few steps in the cycle: nucleotide binding (1-4), SP binding (5-9), GroES binding (10, 11) and, the allosteric transitions that accompany these events (12-17).

The importance of the ADP release phase in the GroEL/GroES catalytic cycle was not appreciated until fairly recently. The results of many early studies were obfuscated by the fact that, when purified from lysed E. coli cells, isolated GroEL contains a diverse array of bound contaminant polypeptides (6, 7). Several newly published studies have demonstrated that the presence of SP in the reaction mixture increases the rate of ADP release, changing the rate limiting step of the cycle to ATP hydrolysis (18-22). This indicates that the ADP release phase is the portion of the cycle that allows the system to be sensitive to cellular demands for folding catalysis.

As the last step in the cycle, ADP release from the GroEL/GroES complex has been largely over-simplified throughout the years. The establishment of the fact that positive cooperativity in ATP binding produced concerted domain movements within a ring (16, 24) ushered in an approach to the GroEL/GroES catalytic cycle that applies the assumption of concertedness to all domain movement dependent steps (15, 23). This is reflected in a number of pre-steady state kinetic

studies that assume ADP release can be modeled as a single step that follows hydrolysis (18, 25, 26).

2.1.2 Overview

This chapter specifically focuses on the developing a kinetic scheme that most accurately describes the release of inorganic phosphate in the catalytic cycle of GroEL. The analysis will help characterize the kinetics of ATP hydrolysis, as well as shed light on a previously underappreciated feature of the ADP release process. The subsequent chapter will build off the foundation laid here, incorporating the results of pre-steady state FRET measurements to further characterize the mechanism of ATP hydrolysis. A background on pre-steady state kinetics is given in Appendix A.2, and the reader is encouraged to read it over so they are familiar with the concepts and terminology used throughout the chapter.

The structure of the analysis presented in the current chapter will be to start with a well rationalized minimal model of the kinetic scheme describing the catalytic cycle of GroEL under the indicated experimental conditions. Conditions will be chosen to minimize the number of states present in the kinetic scheme. Adjustments to the minimal model will be made, when necessary, by pinpointing the specific aspects of the observed data that it fails to accurately describe, and modifying the model accordingly. The goal is to obtain the minimal kinetic scheme necessary to account for the distinctive feature of, and thus best fit, the observed data.

The analysis laid out in the next two chapters provides as-of-yet unforeseen insight into the ATP hydrolysis and ADP release by GroEL. It demonstrates the need for ADP release to be described as a multi-step process. This suggests the importance of the allosteric relaxation ($R \rightarrow T$) in the catalytic cycling of GroEL, which has not been previously demonstrated. Potential insight into this multi-step process will be demonstrated through the molecular dynamics simulations presented in Chapter 5. The analysis also establishes that the kinetics of ATP hydrolysis is well

described as a single step (single exponential) process. These results will be used to determine the mechanism of ATP hydrolysis in the next chapter.

2.2 METHODS

2.2.1 Experimental Methods

In the following analysis, proposed kinetic schemes are tested against a series of rapid mixing experiments. The rapid mixing experiments presented in the following three chapters were performed by my colleague Xiang Ye, and most of the data has been previously published (21, 22). My contribution has been to examine the data in a mathematically rigorous fashion. In doing so, a surprising amount of additional information can be extracted.

The basic setup of all the rapid mixing experiments presented utilizes a stop-flow apparatus. The first syringe contains a sample of GroEL in reaction buffer in one of the two initial state species: the acceptor state complex, [GroES/$GroEL_7$/ADP_7]/[$GroEL_7$], or the apo protein, $GroEL_{14}$. The second syringe contains GroES and ATP. The two are mixed at $t = 0$ with an instrument dead-time of around 4ms.

The release of inorganic phosphate from the complex is detected by including Phosphate Binding Protein (PBP) that is covalently labeled with N-[2-(1-maleimidyl)ethyl]-7-(diethylaminio)-coumarin-3-carboxamide (MDCC) into the reaction buffer (27). The MDCC labeled PBP increases its fluorescence emission upon binding of inorganic phosphate and the binding is both rapid and tight. The increase in fluorescence observed was calibrated using known PBP and known quantities of inorganic phosphate. The measured fluorescence can then be converted into moles of inorganic phosphate released, per mole of $GroEL_{14}$. This is the form of the data presented in the chapter.

2.2.2 Exponential Fitting

As demonstrated in Appendix A.2, a general solution can be obtained for a system of N coupled first order ordinary differential equation's (ODE's) state functions Eq A.2.4 and flux functions A.2.1. In instances where the rate matrix corresponding to the system has a determinant of zero, as is the case in all systems examined in this work, one of its eigenvalues will be zero, further specifying the functional form of the general solution, Eq 2.1. The general solution of a state function, Eq 2.1 can be integrated to give the general solution to the flux of that state, Eq 2.2.

$$X_i(t) = A_1 + \sum_{j=2}^{N} A_{i,j}\, e^{\lambda_j s} \qquad \text{Eq 2.1}$$

$$Y_i(t) = S * t + K + \sum_{j=2}^{N} B_{i,j}\, e^{\lambda_j t} \qquad \text{Eq 2.2}$$

Consider a system with N coupled states and an experiment that monitors the population of a subset of those states. Even if the contribution of each state to observed signal is not known, represented by the arbitrary weighting coefficients $\{c_i\}$, the functional form of the signal will be the sum of N-1 exponentials and a constant. Furthermore, the fitted eigenvalues correspond to the actual rate matrix eigenvalues.

$$S(t) = \sum_{i=1}^{M} c_i\, X_i(t) = \sum_{i=1}^{M} c_i \left(A_{i,1} + \sum_{j=2}^{N} A_{i,j}\, e^{\lambda_j t} \right)$$

$$S(t) = \sum_{i=1}^{M} c_i\ A_{i,1} + \sum_{i=1}^{M} c_i \sum_{j=2}^{N} A_{i,j}\, e^{\lambda_j t}$$

$$S(t) = \sum_{i=1}^{M} c_i\ A_{i,1} + \sum_{j=2}^{N} e^{\lambda_j t} \sum_{i=1}^{M} c_i\, A_{i,j}$$

$$S(t) = \alpha_1 + \sum_{j=2}^{N} e^{\lambda_j t}\ \alpha_j$$

In an exponential fitting, the eigenvalues and coefficients of each of the exponential terms are treated as independent, constant parameters. This is, of course, not true: all coefficients and eigenvalues are functions of the true independent parameters in the system, the micro-constants. Nonetheless, the value in exponential fitting is its ability to establish the minimum number of states required to describe the behavior of the system. A lack of randomness in the fit residuals is a tell-tale sign that too few exponential terms are attempting to be used to describe the behavior of the system.

2.2.3 Reduction of Independent Parameters

The problem with analyzing large systems of coupled equations is that a system of N states has at least N micro-constants that can be independently varied over several orders of magnitude. That implies that the simulation of the full domain of possible micro-constants becomes exponentially more computationally challenging as the number of states increases. Any effort to narrow the domain of possible values must not arbitrarily narrow the system's outcomes.

In the following analysis, portions of the analytical solution to the product accumulation functions will be used in an effort to decrease the number of independent variables. In addition to the sum of exponentials, the general form of the flux function contains two additional terms that describe the steady state behavior of the system. The analytic solution of a given model's product accumulation function will contain a term that is solely a function of the micro-constants and corresponds to the y-intercept of the steady state portion of the graph, $K = K(\{k_i\}_{i=1}^{N})$. The second non-exponential term is linear in time, $S * t$, with $S = S(\{k_i\}_{i=1}^{N})$ also being a function of the micro-constants.

Now assume $\{k_i^*\}_{i=1}^N$ is the set of true micro-constants for a given set of experimental conditions. Then value $K^* = K(\{k_i^*\}_{i=1}^N)$ and $S^* = S(\{k_i^*\}_{i=1}^N)$ would be the y-intercept and slope of the steady state portion of the data obtained under these conditions. Assuming the experiment has been given a sufficient time to reach equilibrium, the value of K^* and S^*can be obtained directly from the data set, by fitting the steady state portion of the graph with a line.

Assume that the functional form of $K(\{k_i\}_{i=1}^N)$ is known from the analytical solution to the kinetic scheme. Then K can be inverted with respect to one of the micro-constants, for example k_J, to give the value of that micro-constant as a function of the remaining micro-constants and K, ie $k_J(\{k_i\}_{i\neq J},\ K)$. Now, the general solution to the chemical master equation of the system, $\vec{X}(t;\ \{k_i\})$, can be rewritten with as:

$$\vec{X}\big(t;\ k_J,\ \{k_i\}_{i\neq J}\big) = \vec{X}\big(t;\ k_J\big(\{k_i\}_{i\neq J}, K\big),\ \{k_i\}_{i\neq J}\big) = \vec{X}(t;\ K,\ \{k_i\}_{i\neq J} \qquad \text{Eq 2.3}$$

The independent variables of the system have been changed, removing one micro-constant, k_J, and replacing it with the term K. However, the true value of the y-intercept, K^*, can be obtained directly from the data set. This value can be plugged into Eq 2.3, reducing the number of independent variables in the equation for the general form of the solution to the proposed kinetic scheme. The same procedure can be performed using the function $S(\{k_i\}_{i=1}^N)$, so the overall number of independent parameters can be reduced by two.

In practice, this procedure is complicated by the fact that the solving either S or K as a function of a micro-constant usually has multiple solutions; the functions $k_J(\{k_i\}_{i\neq J},\ K)$ obtained are rarely proper functions. Thus, the value of $k_J(\{k_i\}_{i\neq J},\ K^*)$ must be checked to ensure it meets the criteria for a reasonable micro-constant, i.e. it is real and positive. Also, it is important to think about how strongly dependent a given term function is on a particular micro-constant. For example, if $k_1 \gg \{k_i\}_{i\neq 1}$, you would not want to invert $S(\{k_i\}_{i=1}^N)$ with respect to k_1 because the steady state rate depends very little on the largest micro-constants.

2.2.4 Bifurcation Analysis

Another technique that will be used to analyze the various kinetic schemes discussed in this chapter is a bifurcation analysis. This type of analysis can be thought of as a brute force method. Bifurcation analysis is the practice of qualitatively analyzing a system of ODEs, under smoothly varying conditions: either initial states or micro-constant values. This involves numerically simulating the results for a given set of micro-constant values, and then repeating the process over a large set of feasible micro-constant values. The results are then plotted against the experimentally obtained data and qualitatively analyzed. In the following analysis, bifurcation analysis will be used to show that, even over a large range of micro-constant values, the simulated curves do not adopt a particular defining feature.

2.2.5 Model Specific Fitting

Unlike exponential fitting, model specific fitting attempts to fit the data obtained in the rapid mixing experiments by varying the system's micro-constants, rather than the macroscopically observable eigenvalues and exponential coefficients. The micro-constants are directly varied within the rate matrix. This means altering a single micro-constant will have corresponding changes in all the system's eigenvalues and exponential coefficients. The results are numerically simulated, because, for all but the smallest systems, this practice is faster than using the analytic form of the solution.

For all fittings, MATLAB was used and its built-in ODE solver ode15s was used to numerically simulate the results. The simulated data compared with the observed data from the rapid mixing experiments and the square of the difference between them is minimized using the MATLAB function *fminsearch*. The results are specific to the proposed kinetic scheme as well as the particular states or fluxes whose corresponding species populations are being observed.

2.2.6 Sensitivity Analysis

The outcome of the model specific fit is a set of values for the various micro-constants, $\{k_i^o\}$, that minimize the difference between the observed data and the function $F(t;\ \{k_i\})$, which is generated from the model. Although its analytical form is never given, the error of the fitted micro-constant values can be analyzed by performing a differential sensitivity analysis on the function $F(t;\ \{k_i\})$. This type of error analysis produces time-dependent sensitivity functions for each independent parameter, $\{\varphi_i(t)\}$, that provide a measure of the sensitivity the output of $F(t;\{k_i\})$ on the input value of the parameter over the domain of the function (28). These sensitivity functions can be plotted with one another to demonstrate the portions of the domain over which each micro-constant has the most influence.

Consider the output function $F(t;a,b)$ and two of its independent parameters a and b, which were fit with values a_o and b_o, respectively. The sensitivity function for the parameter a is given as Eq 2.4 and it will be shown that this produces sensitivity functions that can be compared, on the entire time domain, between micro-constants.

$$\varphi_a(t) = \frac{a_o}{F(t;a_o,b_o)} \left.\frac{\partial F}{\partial a}\right]_{a=a_0} \qquad \text{Eq 2.4}$$

The magnitude of the change of the output function $F(t;a,b)$ that corresponds to the small change in the input value of a is simply the partial derivative of $F(t;a,b)$ with respect to a, $\frac{\partial F}{\partial a}$. Evaluating the partial derivative at the fitted values a_o, b_o gives the change in $F(t;a,b)$ that corresponds to a small change in a about its fitted value, a_o.

The term $\frac{1}{F(t;a_o,b_o)}$ in Eq 2.4 means that the value of the sensitivity function at $t = t^*$ is given relative to the value of $F(t^*;a_o,b_o)$. This allows us to compare the different time points over the course of the experiment. To understand the role of the term a_o in the function, consider the definition of the partial derivative given in Eq 2.5. By evaluating the partial derivative at a given

set of micro-constants we are, by definition, calculating the change in the output function that is associated with a small change in the input parameter.

$$\left.\frac{\partial F}{\partial a}\right]_{a=a_0} = \lim_{\Delta a \to 0} \left.\frac{\Delta F}{\Delta a}\right]_{a=a_0} = \lim_{\Delta a \to 0} \frac{F(t; a_o + \Delta a, b_o) - F(t; a_o, b_o)}{\Delta a}$$

$$\left.\frac{\partial F}{\partial b}\right]_{b=b_0} = \lim_{\Delta b \to 0} \left.\frac{\Delta F}{\Delta b}\right]_{b=b_0} = \lim_{\Delta b \to 0} \frac{F(t; a_o, b_o + \Delta b) - F(t; a_o, b_o)}{\Delta b} \quad \text{Eq 2.5}$$

However, the micro-constants in our model are likely to differ over an order of magnitude or more, so we are not interested in comparing changes to the value of $F(t; a, b)$ that correspond to $\Delta a = \Delta b$. Rather, we are interested in changes to the micro-constants that are proportional to their fitted values. This condition can be written as Eq 2.6. Plugging this condition in to Eq 2.5 and rearranging, you see the term a_o in Eq 2.4 keeps the sensitivity function for each micro-constant proportional to its magnitude.

$$\frac{\Delta a}{a_o} = \frac{\Delta b}{b_o} = c \quad \text{Eq 2.6}$$

$$\left.\frac{\partial F}{\partial a}\right]_{a=a_0} = \lim_{c\, a_o \to 0} \frac{F(t; a_o + c\, a_o, b_o) - F(t; a_o, b_o)}{c\, a_o}$$

$$a_o * \left.\frac{\partial F}{\partial a}\right]_{a=a_0} = \lim_{c \to 0} \frac{F(t; a_o + c\, a_o, b_o) - F(t; a_o, b_o)}{c}$$

As mentioned above, the analytical form of the output functions used in this dissertation are never explicitly given. Their output is usually over the character limit in MATLAB (>25,000) for many of the systems tested and, even when they are obtainable, they are often so large that symbolically evaluating them takes an unreasonably long amount of time. Fortunately, we can evaluate their partial derivative without having to resort to numerical approximations by performing the procedure in Laplace space.

As explained in Appendix A.2.4, the Laplace transform can be used to find the analytical solution to a set of coupled first order ODEs. For most systems, when the micro-constant are symbolically input into the equation, the solution is large if it has an exact solution at all. However, when the value of the input micro-constants is numerically defined, the solution is simply a sum of exponential terms. Thus, if the values of the micro-constants are known, the procedure can be used to generate the desired functions that correspond to the observable in the model.

Fortunately, the following relationship holds for the Laplace transform of partial derivatives (29). If $\mathcal{L}[\ ]$ is the Laplace transform operator and $\mathcal{L}[f(t,x)\ ] = F(s,x)$, then:

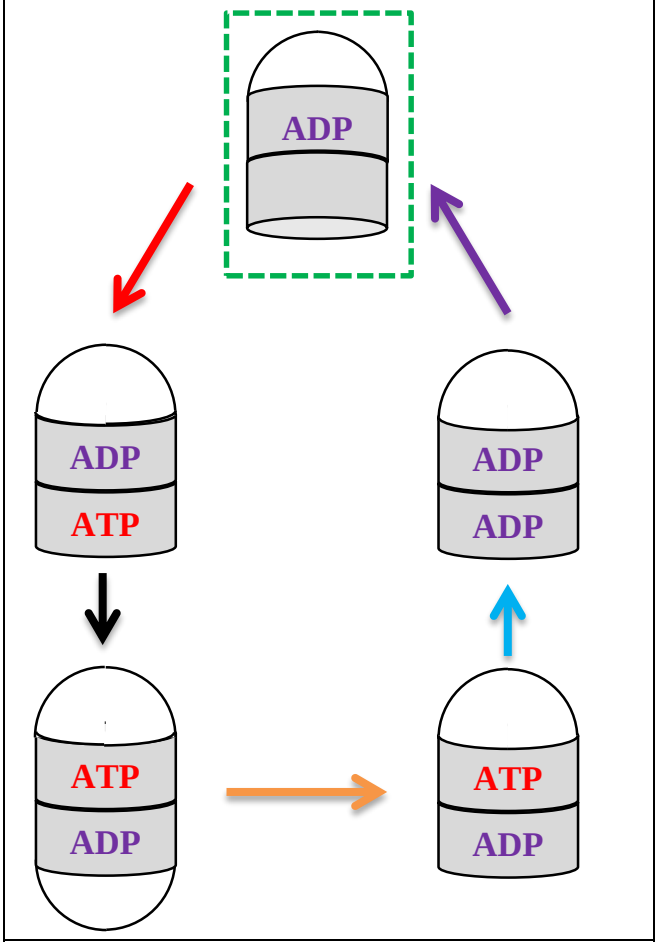
$$\mathcal{L}\left[\frac{\partial f(t,x)}{\partial x}\right] = \int_0^\infty e^{st}\ \frac{\partial f(t,x)}{\partial x}dt = \int_0^\infty \frac{\partial}{\partial x}(e^{st}\ f(t,x))\ dt = \frac{\partial}{\partial x}\int_0^\infty e^{st}\ f(t,x)dt = \frac{\partial F(s,x)}{\partial x}$$

The Laplace transform of the partial derivative of a function is the partial derivative of the Laplace transform of that function. This allows us to perform the scaling analysis by evaluating the the partial derivative of the function of interest in Laplace space, inputting the values of the micro-constants, and then converting the function back to the time domain.

In addition to the sensitivity function, another interesting measure of error in the fitted value of a parameter is the relative sensitivity function, $\psi_i(t)$, defined in Eq 2.7. The sensitivity function gives a measure for how much effect the value of an input parameter has on the value of the output parameter. However, this change, if implemented, may be compensated for by an opposing change in another micro-constant value. The relative sensitivity of a parameter can be thought of as a measure of confidence in the value fitted. A large value of $\psi_i(t)$ at some point in time indicates that a change in k_i could only be compensated for by a much larger change in one of the other micro-constants.

$$\psi_i(t) = \frac{\varphi_i(t)}{\sum_{i=1}^{N}\varphi_i(t)} \qquad \text{Eq 2.7}$$

Figure 2.1 – Starting Catalytic Model for the Asymmetric Cycle of GroEL



Model, obtained from existing literature used as the starting point for building the model used in this chapter. The green box indicates the **initial state** of the system. The arrows represent the catalytic processes: **ATP binding**, **ES binding**, **ES release**, **ATP hydrolysis** and, **ADP release**.

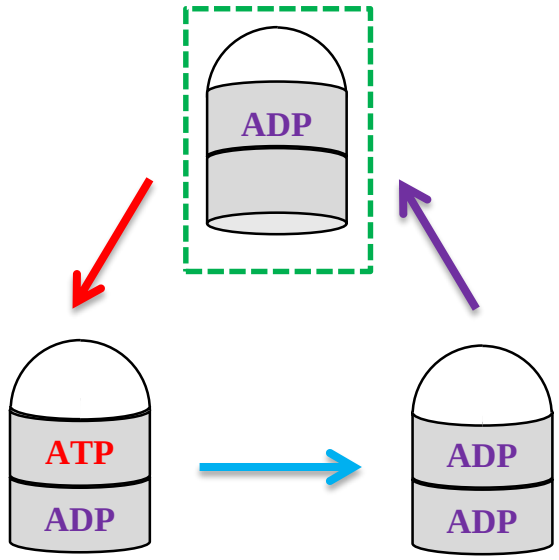
2.3 ANALYSIS

2.3.1 Initial Kinetic Scheme – Three State Model

The starting catalytic model used in the first part of this analysis is shown in Figure 2.1. It contains five distinct species cyclically linked by five kinetic steps. The starting species is the acceptor state complex (green box) which is stable and experimentally isolatable (5, 23). The first step is the binding of ATP to the *trans* ring of the complex. The second step is the binding of GroES, which is closely followed by the release of GroES from the *cis* ring. Next is the hydrolysis of ATP to ADP in the newly formed *cis* ring. Finally ADP is released from the *trans* ring, returning the complex back to its initial state. This is an extremely uncontroversial model for the asymmetric cycle of GroEL, and has been used extensively over the years in a variety of experiments to describe the behavior of the system (20, 22, 30-32).

As presented, the starting kinetic model in Figure 2.1 contains implicit assumptions as to the behavior of subunits within a ring: movement of subunits within a ring is assumed to be concerted. Otherwise, the processes would have to be described as occurring in multiple steps.

Figure 2.2 – Minimal Kinetic Model for the Asymmetric Cycle of GroEL



Simplified model of GroEL catalysis used in this chapter. The green box indicates the **initial state** of the system. The arrows represent the catalytic processes: **ATP/ES binding**, **ATP hydrolysis** and, **ADP release**.

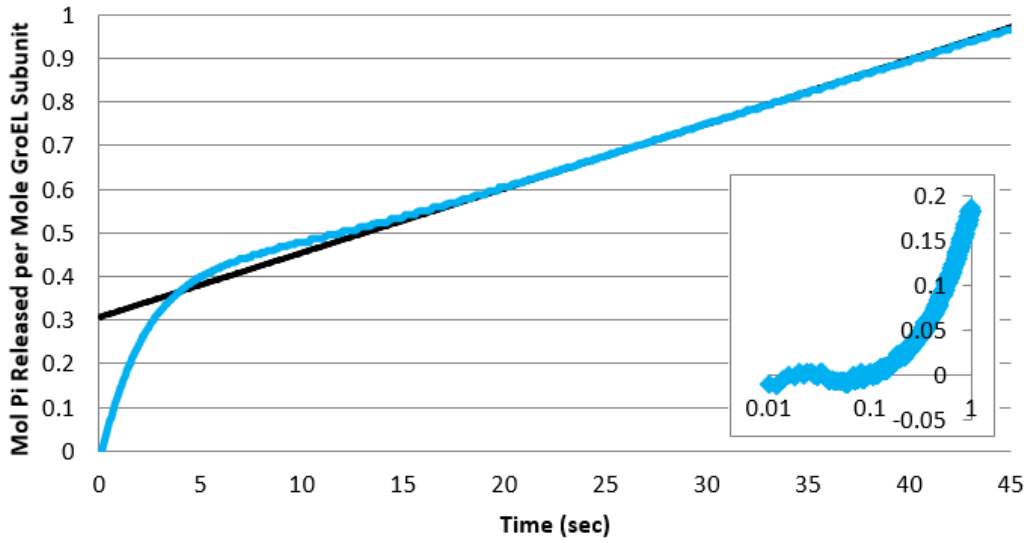
This implies that the processes of hydrolysis and release are both single step, all-or-none processes. The validity of these assumptions is the main focus of the next two chapters.

In order to make the model more manageable, the experimental conditions were chosen in such a way that allows a reduction in the number of species represented in the model. The experiments were performed under conditions where the [ATP], [GroES] and [K^+] were high. When the [ATP] is high, the binding of all 7 molecules of ATP can be approximated by a single step. [GroES] was chosen to be large so that we could combine the first two steps, the binding of ATP and the binding of GroES, into one step. This is, of course, a simplification of reality, but is justified by the magnitude of the binding rates, relative to the rate of hydrolysis, as symbolically justified in Appendix Section A.2.3. [K^+] was kept high so as to slow the release of ADP and minimize the number of species that "leaked" into the symmetric cycle (21, 22). Finally, GroES release from the symmetric $(ADP)_7/(ATP)_7$ complex is also assumed to be so rapid as to be excluded as a state altogether. This approximation is very common in kinetic studies of GroEL (33) and, also, justified symbolically in A.2.3. The minimal kinetic model needed to describe the behavior of the system is then the reduced version of the model in Figure 2.1, represented in Figure 2.2.

2.3.2 Analysis Three State Minimal Model

Our analysis will begin by focusing first on the pre-steady state kinetics of inorganic phosphate release under the conditions laid out above, Figure 2.2. Focusing on this catalytic subsystem reduces the number of states populated throughout the experiment. This, in turn, allows for an analytic solution, where the micro-constants are represented symbolically, to be obtained and analyzed. Even so, the functional form of the solution is still rather large and will not be explicitly presented in the text.

Figure 2.3 – Pre-Steady State Kinetics of Inorganic Phosphate Accumulation in the GroEL/GroES Asymmetric cycle



The **experimental data**, collected by Xiang Ye, obtained by monitoring the accumulation of inorganic phosphate after rapid mixing of GroEL with GroES and ATP. A **best fit line** is fit to the steady-state portion of the data. Notice the portion of the data that lies above the **best fit line**, indicating the convexity of the curve during the delay phase. An inset of the earliest time points, plotted on a log(time) axis, is shown to demonstrate the lag phase of the data.

The data obtained under the specified conditions, collected by Xiang Ye, is presented in Figure 2.3. There is a short lag in the Pi accumulation at the start of the experiment (Figure 2.3, inset), followed by a burst phase that corresponds to the release of roughly 7 molecules of inorganic phosphate per GroEL 14mer complex. The burst phase is followed by a short delay phase, during

which the slope of the graph decreases to below its steady state value, before finally increasing and entering the steady state.

It will now be demonstrated, using a series of analytical techniques, that the three-state minimal model is inadequate to describe the Pi release data obtained. Exponential fitting is used to show that three exponential terms, and thus four states, are necessary to describe the system's behavior. Qualitative analysis of the data's concavity can pinpoint the delay phase as being the problematic region. The three-state model is not able to display the changes in concavity necessary to simulate the data. The subsequent section will focus on the placement of the forth state within the kinetic scheme.

2.3.2.1 Exponential Fitting

Table 2.1 contains the kinetic scheme of the proposed minimal model and, the corresponding chemical master equation and rate matrix. The release of inorganic phosphate is assumed to occur simultaneously with ATP hydrolysis, making its functional form the flux function of state 2, Eq 2.8.

$$\frac{dY_2}{dt} = k_{2,3}X_2 \qquad \text{Eq 2.8}$$

Table 2.1 – Three State Kinetic Scheme

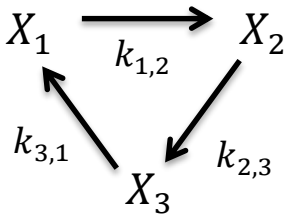
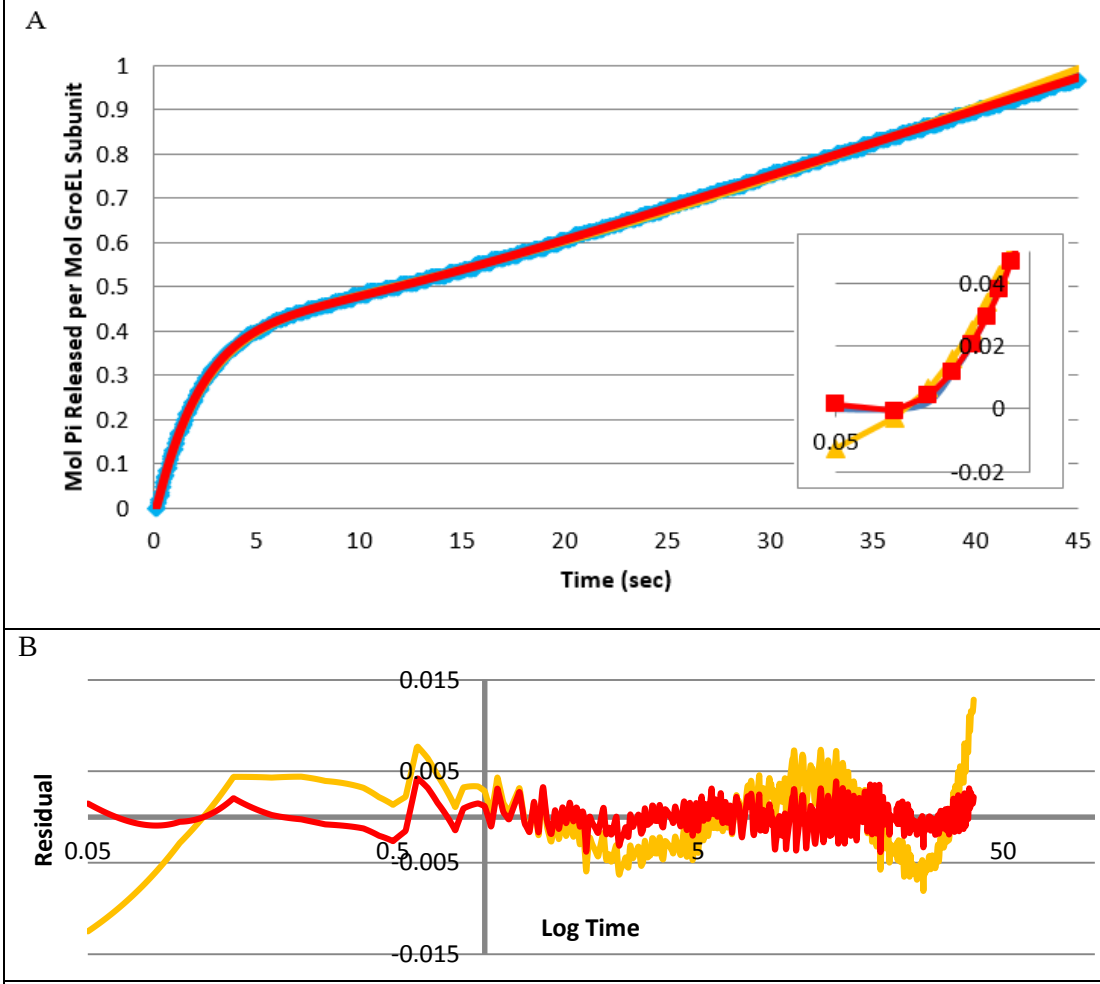
Kinetic Scheme	Chemical Master Equation	Rate Matrix
	$\frac{dX_1}{dt} = -k_{1,2}X_1 + k_{3,1}X_3$ $\frac{dX_2}{dt} = -k_{2,3}X_2 + k_{1,2}X_1$ $\frac{dX_3}{dt} = -k_{3,1}X_3 + k_{2,3}X_2$	$\begin{bmatrix} -k_{1,2} & 0 & k_{3,1} \\ k_{1,2} & -k_{2,3} & 0 \\ 0 & k_{2,3} & -k_{3,1} \end{bmatrix}$

Figure 2.4 – Exponential Fitting of Inorganic Phosphate Accumulation Data to Double and Triple Exponential Functions



(A) The results of an exponential fitting to the **Pi accumulation data**, collected by Xiang Ye. The functions fitted contained a constant, a term linear in time and the sum of **two** or **three** exponentials. The two fits overlay one another almost completely except for the very end and very beginning (inset) of the curve. (B) A plot of the residuals of the **double** and **triple** exponential fittings. The randomness of the residuals of the **triple** exponential fitting demonstrate that it is a better fit than the **double** exponential.

The proposed minimal model contains three conserved states that are cyclically arranged. As highlighted in Appendix A.2.2, in a system of N conserved states, the product accumulation function consists of a sum of N-1 exponential terms, a constant and, a term that is linear in time. An exponential fitting was performed to determine the number of exponential terms, and thus states in the kinetic scheme, necessary to describe the data.

Figure 2.4 shows the results of fitting the data to a double and triple exponential function. For the double exponential, the residuals at the earliest time points show systematic deviation from the best fit curve. Closer inspection of these early time points indicates that the lag phase has been omitted from the fit altogether (Figure 2.4, inset). The fit is thus inconsistent with the proposed model. The addition of a third exponential term successfully accounts for the lag at the beginning of the data set and yielded randomized residuals. This indicates that a minimum of four states are necessary to describe the kinetics of inorganic phosphate release in this system.

2.3.2.2 Concavity Analysis

A qualitative analysis of the concavity of the data set allows us to identify the problematic region. The lag phase is justified by the set of initial conditions. This can be shown by plugging zero into Eq 2.8. The first derivative of the Pi accumulation function is zero at $t = 0$ and the second derivative, equivalent to $\frac{dX_2(0)}{dt}$, is positive. The positive second derivative is consistent with the convexity observed at the beginning portion of the graph.

$$\frac{dY_2(0)}{dt} = k_{2,3}X_2(0) = 0$$

$$\frac{d^2Y_2(0)}{dt^2} = k_{2,3}\frac{dX_2(0)}{dt} = k_{2,3}\left(-k_{2,3}X_2(0) + k_{1,2}X_1(0)\right) = k_{2,3}k_{1,2}$$

The burst phase is a widely observed phenomenon in pre-steady state kinetics, and occurs when the observed product release step in the cycle (here, Pi release upon ATP hydrolysis) is followed by a slower step in the cycle (here, the release of the other product ADP), which decreases the overall steady-state rate of product release. During this phase the concavity changes, indicating the second derivative of the flux function corresponding to the product accumulation curve has a zero in this region, changing sign from positive to negative. The graph then changes back to being convex during the delay phase. This is especially apparent when a straight line is fit to the linear

portion of the data, demonstrating the convexity in this portion of the graph, Figure 2.3. This indicates that the second derivative of the flux function corresponding to the product accumulation curve has a zero in this region as well, changing sign from negative to positive.

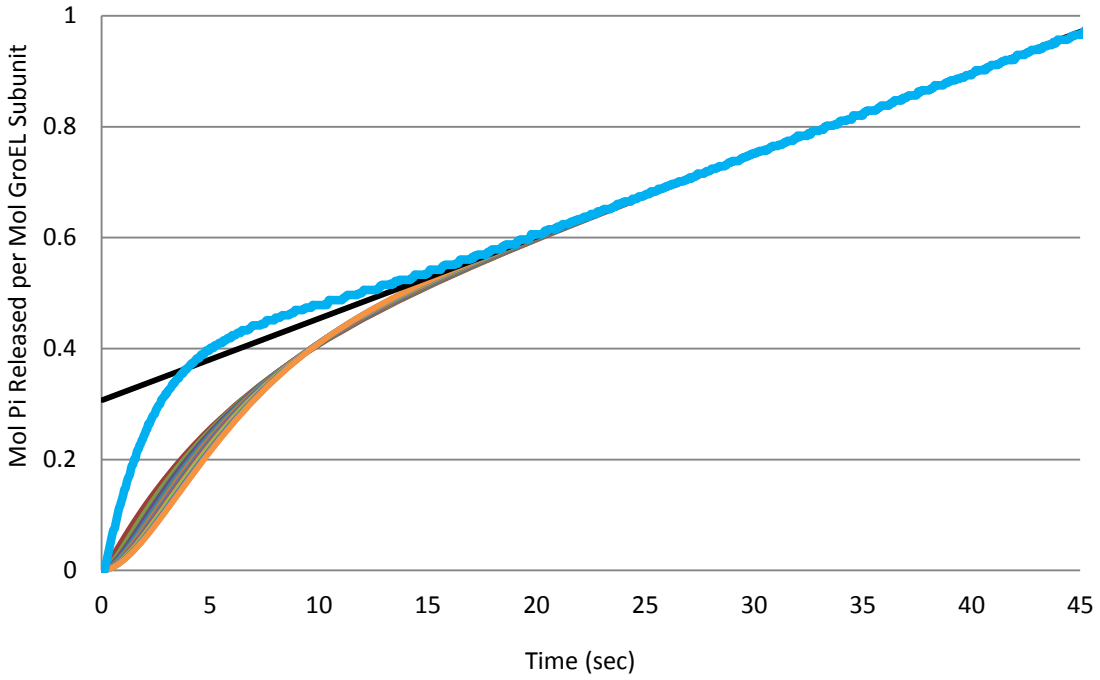
The second change in the data curve's concavity is inconsistent with the three state minimal model proposed. Two changes in concavity indicate that the second derivative of the curve must have two zeros. The second derivative of the flux function of a three state system consists of a sum of just two exponential terms. It is not difficult to show that a function consisting of the sum of N exponentials with real rate terms has at most N-1 zeros (34). Thus, a sum of two exponential terms can only be equal to zero at one point in its domain.

2.3.2.3 Bifurcation Analysis

The above argument breaks down if the eigenvalues of the rate matrix have non-zero imaginary parts. Although this is unlikely given the difference in magnitude of the fitted eigenvalues (see Appendix A.2.5), we can qualitatively demonstrate the inability of the three-state model to account for the delay phase using an exhaustive bifurcation analysis that accounts for the possibility of imaginary eigenvalues. The number of independent parameters of the system was reduced, as described in Section 2.2.3 above, using the fitted values for the slope and y-intercept of the steady state portion of the curve. The remaining micro-constant was varied over six orders of magnitude, from 10^{-3} to 10^{3}, in increments equally spaced in log-space. The values that produced realistic values for the other two micro-constants (i.e. real and positive) were plugged into the rate matrix for the system and solved numerically.

An example set of curves from the bifurcation analysis that cover the full ranges of observed errors is given in Figure 2.5. The curves are forced to have the correct steady state portion of the curve due to the reduction of independent parameters. Nonetheless, none of the combinations of micro-constants are able to accurately simulate the data. This is a result of the inability of the

Figure 2.5 – Bifurcation Analysis of the Three State System



The results of the bifurcation analysis of the kinetic scheme presented in Table 2.1. The experimental **Pi accumulation data**, collected by Xiang Ye, is plotted alongside the family of curves simulated during the **bifurcation analysis** (all other curves plotted). A **best fit line** is provided to demonstrate that none of the simulated curves demonstrate the changes in concavity observed in the data. This supports the conclusion that the three state model is not sufficient to describe the observed data.

model to accurately account for the delay phase of the curve. This is indicated by the fact that none of the combinations of micro-constants used produced curves that lay above the steady-state line (Figure 2.5, black).

2.3.2.4 Summary of Results

The fitting of the data to the general solution of the product accumulation function for the proposed minimal kinetic scheme implied that the model lacked the proper number of states. This was further supported by qualitative analysis of the concavity of the observed data set, and comparing it to the predicted model. Finally, bifurcation analysis was used to demonstrate that, even if the eigenvalues of the rate matrix had imaginary components, they could not accurately simulate the observed changes in concavity. Thus, the model is insufficient to describe the system's behavior due to the delay phase.

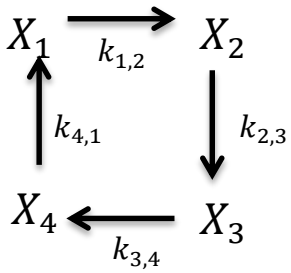
2.3.3 Four State Model

Analysis of the three state model has allowed us to conclude that no choice of values for the micro-constants will produce the concavity profile observed in the data set. The model thus needs to be expanded to include a fourth state. We will begin by doing so from a strictly mathematical perspective; not bothering *a priori* what species is represented by the state. This analysis will allow us to determine, relative to the Pi release state, where the fourth state must be placed to yield simulated data with the correct concavity profile.

The fact that an additional state is needed to accurately characterize the system calls into question the validity of the numerous assumptions made when constructing the three-state model, which limited the number of states (from Figure 2.1 to Figure 2.2). It will be demonstrated that relaxing only one of the various assumptions implemented can reconcile the observed discrepancy.

2.3.3.1 Analysis of Four State Model

Table 2.2 – Four state kinetic scheme

Kinetic Scheme	Chemical Master Equation	Rate Matrix
	$\frac{dX_1}{dt} = -k_{1,2}X_1 + k_{4,1}X_4$ $\frac{dX_2}{dt} = -k_{2,3}X_2 + k_{1,2}X_1$ $\frac{dX_3}{dt} = -k_{3,4}X_3 + k_{2,3}X_2$ $\frac{dX_4}{dt} = -k_{4,1}X_4 + k_{3,4}X_3$	$\begin{bmatrix} -k_{1,2} & 0 & 0 & k_{4,1} \\ k_{1,2} & -k_{2,3} & 0 & 0 \\ 0 & k_{2,3} & -k_{3,4} & 0 \\ 0 & 0 & k_{3,4} & -k_{4,1} \end{bmatrix}$

The exponential fit, Figure 2.4, indicates that three exponential terms satisfactorily fits the data, indicating a minimum of four states may be sufficient to describe the system's behavior. Assuming that the correct model contains four coupled states, where in the cycle should the new

state be placed? A mathematically equivalent way to formulate this question is to start with a four state kinetic scheme and compare the various flux functions, $\{Y_i(t)\}_{i=1}^{4}$, of each state in order to determine which one best simulates the data. Table 2.2 contains the kinetic scheme of the four state model proposed, along with its corresponding chemical master equation and rate matrix.

The analytic solution to $Y_4(t)$ can be obtained by using the inverse Laplace transform method outlined in Appendix A.2.4. The equation for the y-intercept for this particular flux function is given as Eq 2.9. In order to make the equation more easily readable, the set of variables $\{a, b, c, d\}$ is used in place of $\{k_{1,2}, k_{2,3}, k_{3,4}, k_{4,1}\}$. Since the value of a micro-constants is always positive, the y-intercept term of the flux function $Y_4(t)$ is non-positive for all possible combinations of micro-constants. This rules out the possibility that the fourth step is the Pi release step.

$$K_4 = \frac{-7(ad + bd + cd + ac + bc + ab)(abcd)}{(bac + bda + acd + bcd)^2} \qquad \text{Eq 2.9}$$

The general form of the flux function $Y_i(t)$, Eq A.2.1, can be differentiated to obtain the general form for its first and second derivative, Eq 2.10 and Eq 2.11, respectively.

$$Y_i(t) = \int_0^t k_i X_i(s)\, ds$$

$$\frac{dY_i(t)}{dt} = k_i X_i(t) \qquad \text{Eq 2.10}$$

$$\frac{d^2 Y_i(t)}{dt^2} = k_i \frac{dX_i}{dt}(t) \qquad \text{Eq 2.11}$$

For the flux function of the first state of the four state model (Table 2.2), $Y_1(t)$, plugging in the initial conditions into Eq 2.11 demonstrates that the second derivative of the function is negative at $t = 0$. For that to be true, the curve must be concave at the beginning portion of the graph,

which is not consistent by the qualitative analysis performed on the concavity of the data. This rules out the first step as being the Pi release step.

$$\frac{d^2Y_1(t)}{dt^2} = k_1\frac{dX_1}{dt}(t) = k_1(-k_1X_1(t) + k_4X_4(t))$$

$$\frac{d^2Y_1(0)}{dt^2} = k_1(-k_1X_1(0) + k_4X_4(0)) = -{k_1}^2$$

For the flux function of the third state, $Y_3(t)$, plugging in the initial conditions into Eq 2.6 demonstrates that the second derivative of the function is zero at $t = 0$. The concavity analysis indicated that the data changes concavity twice throughout the course of the experiment. However, the general form of the second derivative of the flux function $Y_3(t)$ is a sum of three exponential terms. In the same argument as was made above, a sum of three exponentials with real rate terms has, at most, two zeros. If one of them is located at $t = 0$, then on $t \in (0, \infty)$, there can be at most one other zero and thus only one change in concavity on this domain.

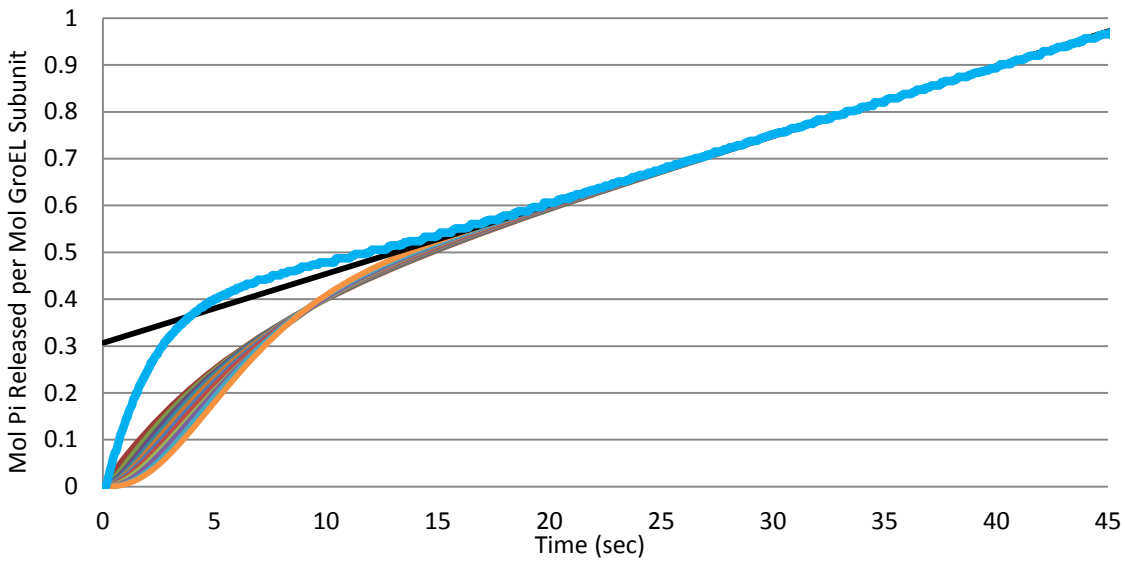
$$\frac{d^2Y_3(t)}{dt^2} = k_3\frac{dX_3}{dt}(t) = k_3(-k_3X_3(t) + k_2X_2(t))$$

$$\frac{d^2Y_3(0)}{dt^2} = k_3(-k_3X_3(0) + k_2X_2(0)) = 0$$

2.3.3.2 Bifurcation Analysis

The above argument breaks down if the eigenvalues of the system have non-zero complex parts. In order to conclusively demonstrate that the third step is not the Pi release step, a bifurcation analysis was performed similar to the one described for Figure 2.5. Again, the number of independent parameters was reduced from four to two using the procedure outlined above. The remaining two micro-constants, $k_{1,2}$ and $k_{2,3}$, were varied from 10^{-3} to 10^3 in 100 increments equally spaced in log-space.

Figure 2.6 – Bifurcation Analysis of Product Release from State 3 of Four State System



The results of the bifurcation analysis of the kinetic scheme presented in Table 2.2 with product release occurring at the third step. The experimental **Pi accumulation data**, collected by Xiang Ye, is plotted alongside the family of curves simulated during the **bifurcation analysis** (all other curves plotted). A **best fit line** is provided to demonstrate that none of the simulated curves demonstrate the changes in concavity observed in the data. This supports the conclusion that hydrolysis occurs at the second step.

A representative set of curves obtained from the bifurcation analysis, which cover the full range of observed errors, is plotted in Figure 2.6. Again, the curves are forced through the steady state portion of the graph due to the reduction of independent parameters. Nonetheless, none of the combinations of micro-constants are able to correctly simulate the observed data. The error can be attributed to the fact that none of the generated curves of the flux function $Y_3(t)$ have the necessary concavity profile, and thus are unable to simulate the observed delay phase. This leads to the conclusion that third step in the kinetic scheme is not the Pi release step.

2.3.3.3 Model Specific Fits

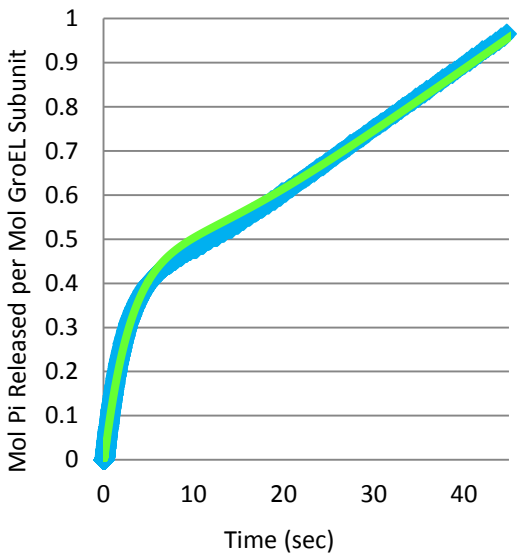
The only member remaining in the set of possible flux functions for the system is $Y_2(t)$ and it cannot be disqualified using any of the arguments made above. The data was fit to $Y_2(t)$ using a model specific fit. The results are shown in Figure 2.7. The flux function $Y_2(t)$ is able to simulate the data very well as indicated by the high coefficient of determination. The fit improves ten-fold

when the potassium concentration is increased to 200mM K+. Under these conditions, the release of ADP is even slower, preventing the system from leaking into the symmetric cycle. Despite the goodness of fit and its ability to simulate all the observed phases of the data, the residuals are not completely random, indicating the kinetic scheme is only an approximation.

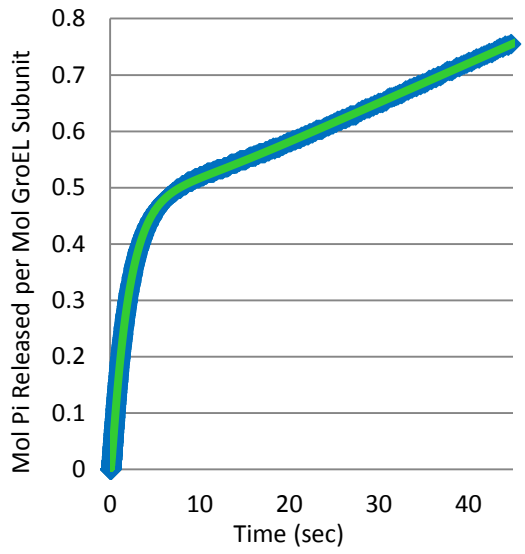
Figure 2.7 – Results of Model Specific Fit of Product Release from State 2 of a Four State Model

100mM K^+	200mM K^+

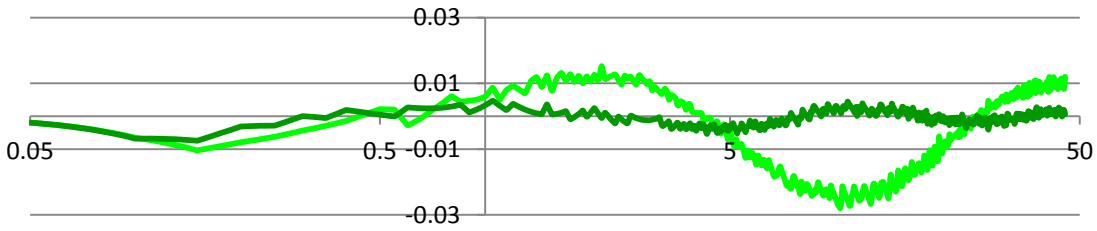
A



B



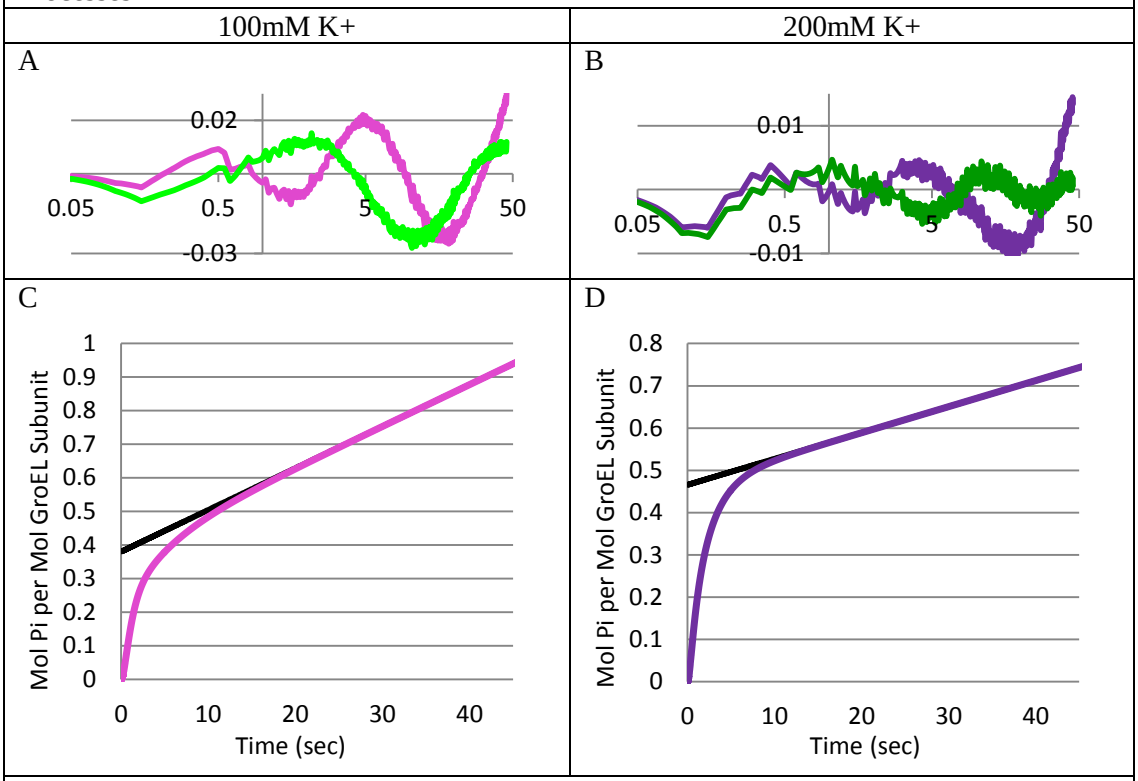
C



Fitted Micro-constant Values (sec^{-1})			
100mM K^+		200mM K^+	
k_1	11.1	k_1	6.89
k_2	0.33	k_2	0.505
k_3	0.063	k_3	0.016
k_4	0.063	k_4	0.21
R^2	0.9983	R^2	0.9998

Model specific fit of the kinetic scheme outlined in Table 2.2 with product release occurring at State 2 to the Pi accumulation data. (A) **Best fit curve** overlaid on the experimental **Pi accumulation data**, collected by Xiang Ye, obtained in buffer with **100mM K^+**. (B) **Best fit curve** overlaid on the experimental **Pi accumulation data**, collected by Xiang Ye, obtained in buffer with **200mM K^+**. (C) The residuals of the model specific fits at **100mM K^+** and **200mM K^+** plotted on a log time scale.

Figure 2.8 – Results of Model Specific Fit Comparing Single-Step and Two-Step Hydrolysis Processes



Model specific fits comparing single-step and two-step hydrolysis in the kinetic scheme outlined in Table 2.2. (A) The residuals of the fit to the data collected at 100mM K^+ of the **single-step hydrolysis** model (Figure 2.7.A) and the **two-step hydrolysis model**. (B) The residuals of the fit to the data collected at 200mM K^+ of the **single-step hydrolysis** model (Figure 2.7.B) and the **two-step hydrolysis** model. (C) and (D) Plots of the best fit curve simulated using the two-step hydrolysis model at **100mM K^+** and **200mM K^+**, along with the **best fit line** to the linear portion of the simulated data, demonstrating the absence of the delay phase in the two-step hydrolysis model.

Thus far, it has been demonstrated that a fourth state is necessary to accurately describe the behavior of the system. The exponential fit indicated that a four state model could be sufficient to describe the behavior of the system. The model specific fit of Figure 2.7 demonstrated that the addition of another state after the hydrolysis step drastically improved the goodness of fit and successfully accounted for the expected changes in concavity. The kinetic scheme tested in Figure 2.7 followed the initially stated assumption that ATP hydrolysis, and thus Pi release, occurs as a single step. In order to test the validity of this assumption, we will perform a model specific fit using a flux function that is a combination of $Y_2(t)$ and $Y_3(t)$. This would correspond to a system with a two-step hydrolysis process. If the true hydrolysis process is a series of two or more steps, the fit to the new model should be an improvement.

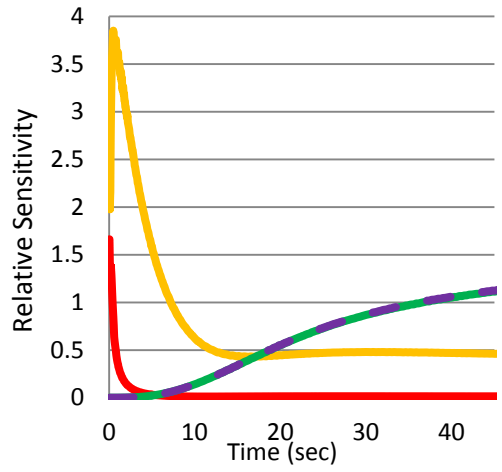
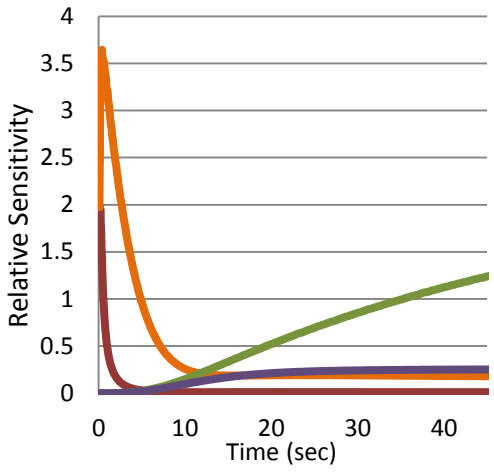
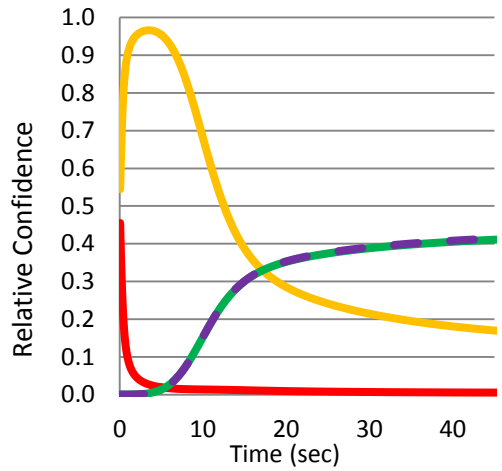
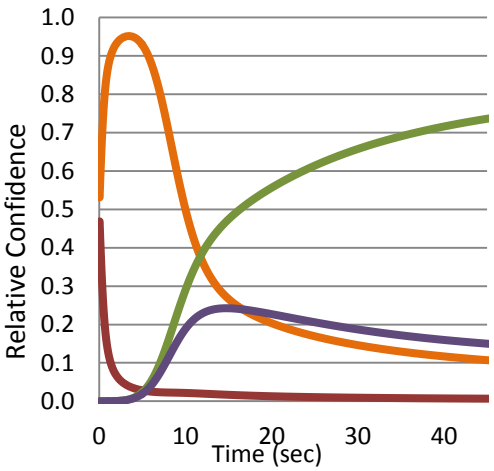
The residuals of a model specific fit to the two step hydrolysis process are shown in Figure 2.8.A and 2.8.B. In both cases, relaxing the assumption of single step ATP hydrolysis does not improve the fit. The error in the two fits is only slightly greater, but in both cases the residuals are even less randomly distributed and the delay phase is completely lost. This is a result of the fact that the height of the burst phase forces the third micro-constant to be so high as to lose the key delay phase characteristic in the data. The loss of the delay phase is most evident when the steady-state portion of the simulated curves is fit with a line Figure 2.8.C and 2.8.D.

2.3.3.4 Sensitivity Analysis

In order to determine the error associated with the fitted values of each of our micro-constants, a sensitivity analysis was performed on the system using the procedure outlined in Section 2.2.6. The usual reporting of standard error of a variable would insufficiently describe the error of a fitted micro-constant value due to its variation in time. Changes to different micro-constant values affect different regions of the output function by many orders of magnitude. Thus, a time dependent description of error is better suited for the system. Figure 2.9 are plots of the sensitivity functions (A and C) and confidence functions (B and D) obtained by applying Eq 2.4 and 2.7 to the flux function of the second state in the four state model of Table 2.2. The fitted values used in the analysis are given in Figure 2.7.

The sensitivity analysis demonstrates how differently the various micro-constants affect the output of the system at different points in time. At time values close to zero, the system is extremely sensitive to the first micro-constant, and practically independent thereafter. The third and fourth micro-constants have identical sensitivity functions for the 100mM K^+ conditions due to the fact that they were fitted to the same value. For the 200mMK^+ the third micro-constant has a greater impact on the steady-state portion of the graph than the fourth, due to the fact that it is the smallest of the four micro-constants.

Figure 2.9 – Sensitivity and Confidence Analysis of the Model Specific Fits in Figure 2.7

100mM K^+	200mM K^+
Sensitivity Functions	
A 	C 
Confidence Functions	
B 	D 
Sensitivity (A) and confidence (B) functions for the fitted values, obtained at 100mM K^+, in Figure 2.7 of k_1 , k_2 , k_3 and, k_4. The functions corresponding to k_3 are dotted to make those of k_4 visible.	Sensitivity (C) and confidence (D) functions for the fitted values obtained at 200mM K^+, in Figure 2.7 of k_1 , k_2 , k_3 and, k_4.

The second micro-constant has the largest impact over the region in time corresponding to the burst phase. Under both potassium concentrations, the output functions are extremely sensitive to the second micro-constant before the system enters the steady state. The confidence analysis of the two systems demonstrates that the second micro-constant is almost solely responsible for the simulated behavior of the system in the first ~7 seconds of the curve. The position of the delay

phase coincides with the rise in influence of both the final two micro-constants, indicating the importance of their value (i.e. magnitude) in producing the concavity changes observed.

2.3.3.5 Justification of Initial Assumptions

At the beginning of the chapter, many assumptions were made about the catalytic mechanism of GroEL that allowed for a simplified kinetic scheme with a reduced number of states. The assumptions that binding of all 7 ATP occurred in a single step, ATP and GroES binding occurred simultaneously, and GroES release from the symmetric $(ADP)_7/(ATP)_7$ complex was so rapid that the state could be omitted altogether, reduced the first portion of the kinetic scheme to a single step. If any of these assumptions led to the failure of the initial three-state kinetic scheme, Table 2.1, then the addition of a state between states 1 and 2 should have rectified the problem. By demonstrating that the flux function of the third state was unable to simulate the observed concavity, specifically the important delay feature of the data, the notion that relaxing any one of these three assumptions would allow us to more accurately simulate the data is undermined.

2.3.3.6 Summary of Results

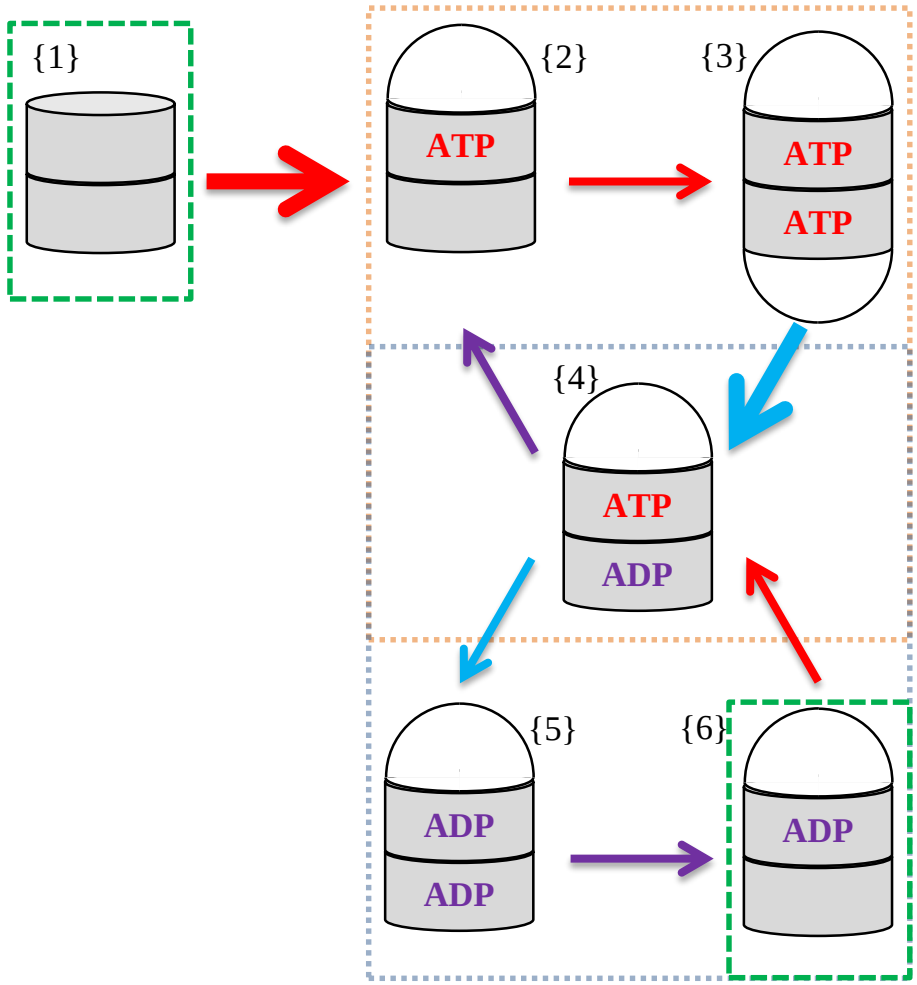
Analysis of the pre-steady state kinetics of inorganic phosphate release of the asymmetric complex has helped to build upon the initial catalytic model proposed in Figure 2.2. The assumption that ATP and GroES binding occur simultaneously and within a single step was shown to be reasonable under the experimental conditions chosen. It is also reasonable to omit the symmetric $(ADP)_7/(ATP)_7$ complex altogether, because GroES dissociation is so rapid from this state. Lastly, and most importantly, the assumption that Pi release occurs as a single step was shown to hold.

Contrary to our initial assumption, the preceding investigation demonstrated that ADP release cannot be accurately depicted as a single step process. The concavity profile of the data can only be accurately simulated with the placement of an additional state in this portion of the kinetic

scheme. Doing so leads to the minimal model that can reliably simulate the data (Figure 2.7). The possible nature of the species corresponding to the additional forth state to the cycle is yet unknown and will be discussed at the end of the chapter.

2.3.4 Expanding Analysis to Apo-GroEL Initial State

Figure 2.10 – The Minimal Kinetic Model for the Full Catalytic Cycle of GroEL



Simplified kinetic model of the full GroEL catalytic cycle. The green box indicates the **initial states** of the system. The arrows represent the catalytic processes: **ATP/ES binding**, **ATP hydrolysis** and, **ADP release**. The orange dotted box encloses the **symmetric cycle** and the blue dotted box encloses the states of the **asymmetric cycle**. The thickened arrows indicate that either ring can undergo the indicated process and the results would be indistinguishable, so the rate constant at this step must be double its normal value.

The focus will now be on an expanded version of the initial catalytic model proposed in Figure 2.2, shown in Figure 2.10. This catalytic model describes the behavior of the GroEL/GroES

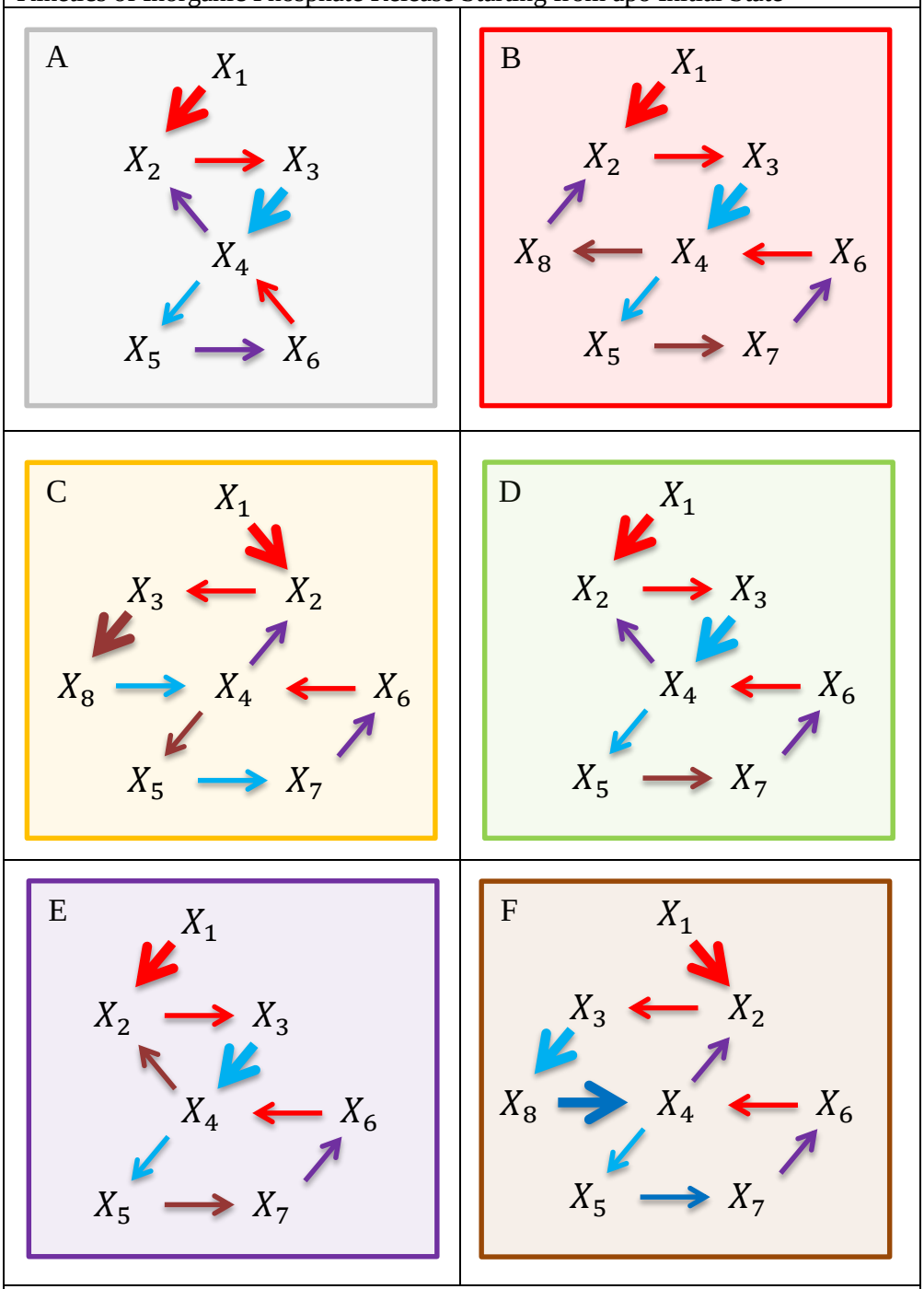
system when the apo state is the initial state of the system. The model was previously proposed because it successfully accounts for the behavior of the system in the absence and presence of SP (20-22). The model presented in Figure 2.10 is a simplified version of this published model, with the same assumptions applied that were used in Section 2.3.1 to reduce the catalytic model in Figure 2.1 to the one used in the above analysis, Figure 2.2.

When starting with apo-GroEL as the initial state, the system accesses many more states, and the analytic solution to the system of equations is no longer reasonable to work with. In order to investigate the catalytic model proposed in Figure 2.10, we will have to rely much more heavily on numerical analysis. In order to keep the number of independent parameters to a minimum, all binding events will be assumed to occur at the same rate; the same holds true for all other like micro-constants. Because of the indistinguishability of the two rings, as drawn, some states actually represent two distinct species (i.e. State 2 of Figure 2.10 is the sum of species with ATP/GroES bound to ring A and those with ATP/GroES bound to ring B). In such situations, the rate constants are modified in order to reflect this truth (i.e. if k_B is the binding micro-constant, than $2k_B$ is the rate of the process connecting State 1 and State 2 and k_B is the rate of the process connecting State 2 and State 3).

2.3.4.1 Kinetic Schemes

The previous analysis demonstrated that the kinetic scheme obtained from the asymmetric cycle portion of Figure 2.10 (blue dotted box) was inadequate to describe the behavior of the system. An additional state, inserted in the portion of the cycle corresponding to the release of ADP, was required in order to correctly describe the pre-steady state behavior of the asymmetric cycle. To address whether a similar modification to the kinetic scheme of the catalytic model in Figure 2.10 is necessary, the series of closely related kinetic schemes shown in Figure 2.11 are compared using model specific fits.

Figure 2.11 – Set of Kinetic Schemes Compared in Modeling of Pre-Steady State Kinetics of Inorganic Phosphate Release Starting from apo Initial State



The various kinetic schemes compared using model specific fits, Figure 2.10. The arrows represent the catalytic processes: **ATP/ES binding**, **ATP hydrolysis**, **ATP hydrolysis 2** (F only), **ADP release 1**or **random** (C only) and, **ADP release 2**. The thickened arrows indicate that either ring can undergo the indicated process and the results would be indistinguishable, so the rate constant at this step must be double its normal value. Background colors are consistent with corresponding fit curve colors in Figure 2.10.

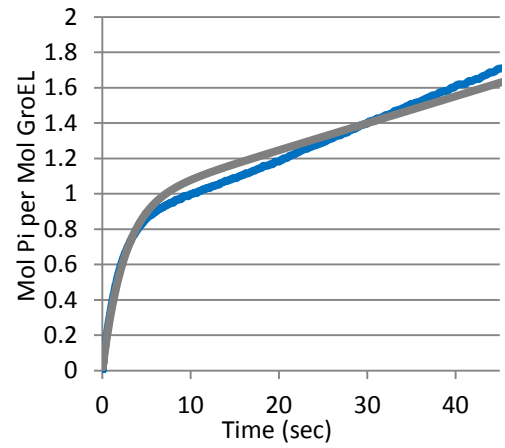
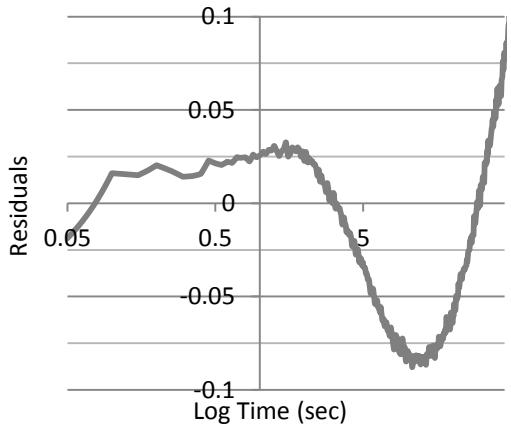
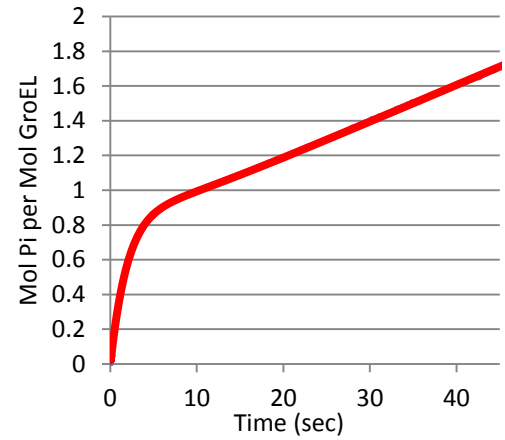
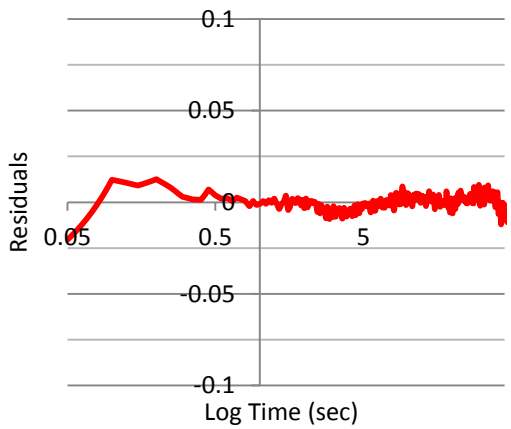
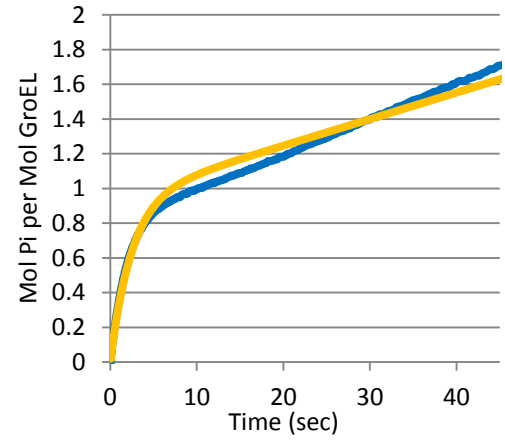
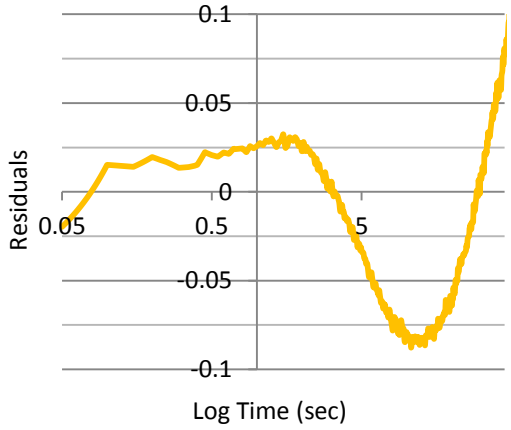
The first of the six kinetic schemes (Figure 2.11.A) is simply the kinetic scheme of the initial catalytic model presented in Figure 2.10, which describes ADP release as a single step. The second (Figure 2.11.B) has two additional states, each inserted in one of the two ADP release portions of the cycle. The addition of the new process in Figure 2.11.B introduces a new micro-constant, and thus independent parameter, into the fitting procedure, which will inevitably increase the flexibility of the model. The third kinetic scheme (Figure 2.11.C) is included in the analysis to test whether or not an additional independent parameter anywhere in the cycle is sufficient to reproduce the observed data. This third scheme also has two additional states compared to the initial kinetic scheme of Figure 2.11.A, but at different positions in the cycle than the scheme in Figure 2.11.B.

The fourth and fifth models (Figure 2.11.D and 2.11.E, respectively) are included in the analysis to test the importance of the additional state in the ADP release portion of symmetric cycle. The previous analysis established the importance of this state in the asymmetric cycle, but this does not necessarily imply the same is true for the symmetric cycle. The last kinetic scheme (Figure 2.11.F) tested contains a two-step hydrolysis process. As was done previously in Section 2.3.3.3, this scheme is included in order to ensure that the assumption of single-step hydrolysis is not preventing the model from accurately simulating the observed data.

2.3.4.1 Model Specific Fits

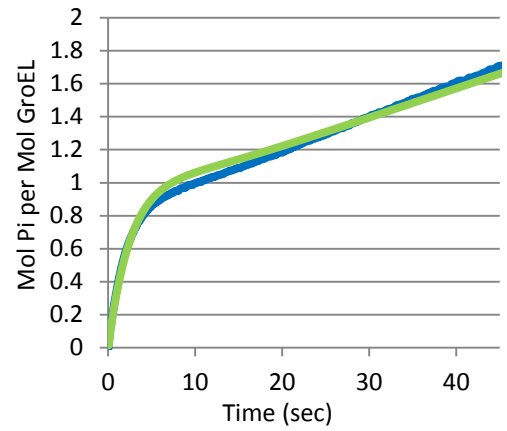
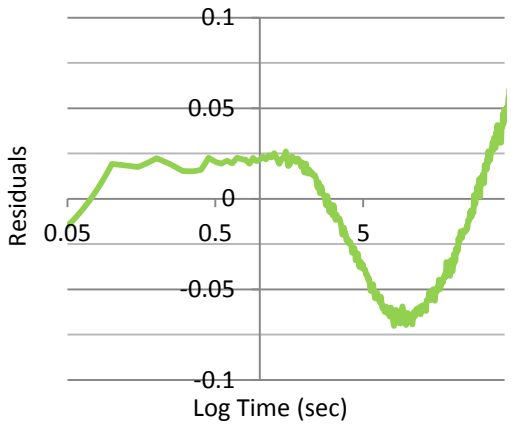
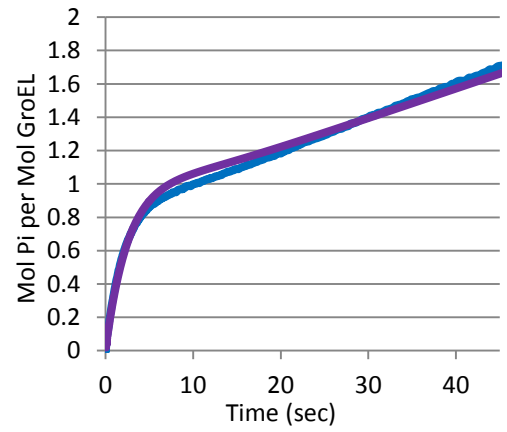
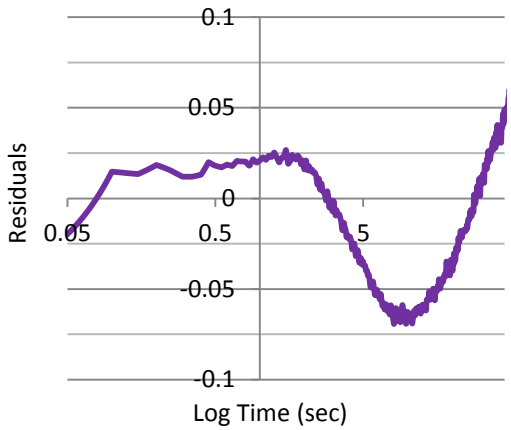
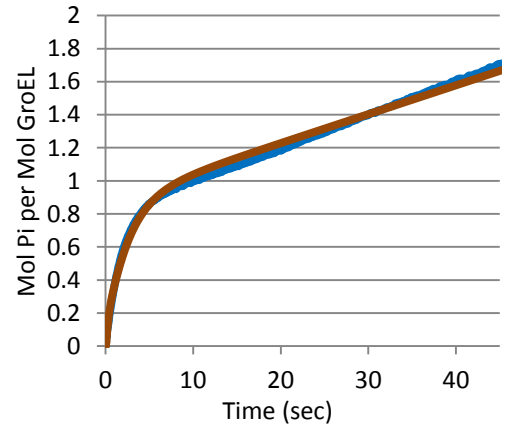
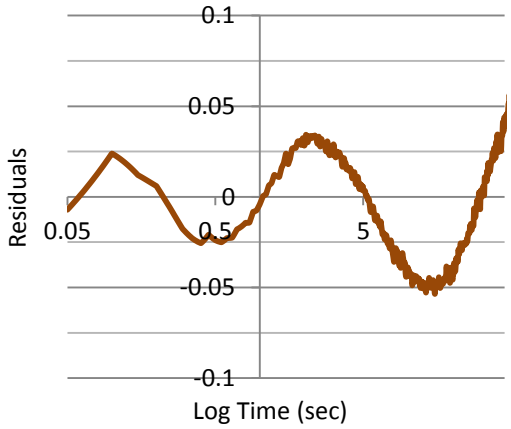
The six kinetic schemes outlined in Figure 2.11 were subject to model specific fits. The results are presented in Figure 2.12 and Table 2.3. The one-step ADP release of the first scheme is not able to successfully capture the observed delay phase. The third scheme also fails to capture the delay phase, indicating that the introduction of a new independent parameter was not sufficient to improve the fit. The fourth and fifth schemes demonstrate the importance of representing both ADP release portions of the cycle as multi-step. Finally, the two-step hydrolysis scheme,

Figure 2.12 – Model Specific Fits of Various Kinetic Schemes Outlined in Figure 2.9 to Inorganic Phosphate Accumulation Data Starting with Apo-GroEL as the Initial State

Model Specific Fit	Fit Residuals
A.1 	A.2 
B.1 	B.2 
C.1 	C.2 

Model specific fits of the **Pi accumulation data**, obtained by Xiang Ye, to the various kinetic schemes of Figure 2.9. Panel 1 (left) contains the fitted curve overlaid on the **data** while Panel 2 (right) is the residuals of the fit. Lettering and coloring is consistent with Figure 2.9

Figure 2.12 Cont – Model Specific Fits of Various Kinetic Schemes Outlined in Figure 2.9 to Inorganic Phosphate Accumulation Data Starting with Apo GroEL as the Initial State

Model Specific Fits	Fit Residuals
D.1 	D.2 
E.1 	E.2 
F.1 	F.2 

Model specific fits of the **Pi accumulation data**, obtained by Xiang Ye, to the various kinetic schemes of Figure 2.9. Panel 1 (left) contains the fitted curve overlaid on the **data** while Panel 2 (right) is the residuals of the fit. Lettering and coloring is consistent with Figure 2.11

presented in scheme six, showed that the addition of a second hydrolysis step does not improve the model's ability to describe the data. The second scheme, in which both ADP release portions of the cycle are considered to be a two-step process, is able to accurately simulate the data. The high coefficient of determination and randomness in the fit residuals indicate the success of the simplified scheme in Figure 2.11.B.

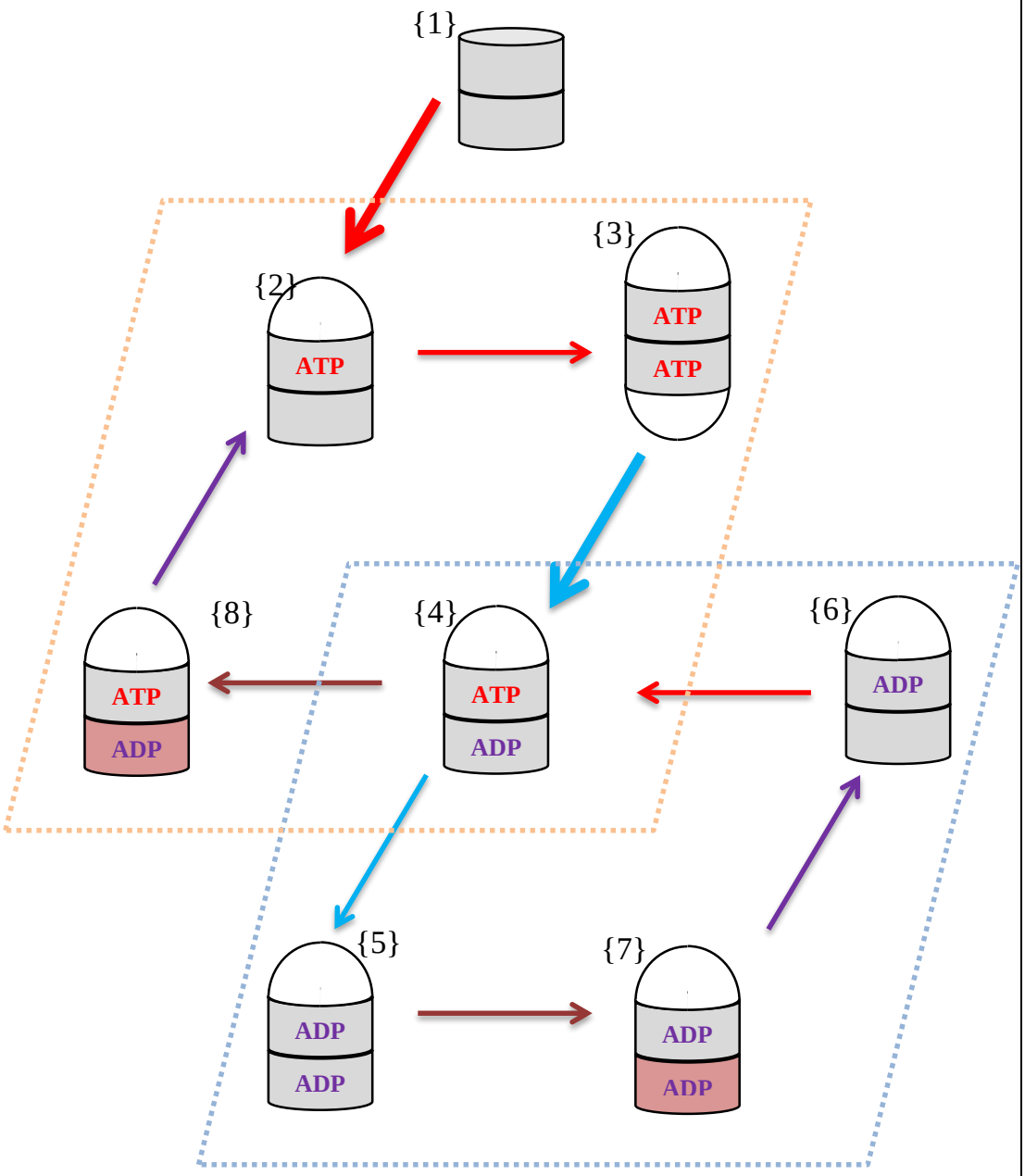
2.4 CONCLUSION

Our pre-steady state investigation of the kinetics of inorganic phosphate release from the GroEL/GroES system has allowed us to build a minimal kinetic scheme that accurately describes the behavior of the system under the conditions chosen, Figure 2.13. ATP and GroES binding are approximated as a single step in the model, which is sufficient to describe the data obtained due to the high concentration of each species used. The species corresponding to the symmetric $(ADP)_7/(ATP)_7$ complex can be omitted from the kinetic scheme altogether due to the rapid rate of GroES release. These assumptions were validated in Section 2.3.3 by comparing the flux function of each state and discarding the ones that do not have the correct concavity signature.

ATP hydrolysis was assumed to be a single step, accompanied by the release of inorganic phosphate. It was shown that relaxing this assumption does not improve the fit in any of the data sets collected. The overall goodness of fit to the final models indicates that the process of hydrolysis in the systems can be well approximated by a single step. This result will be used extensively in the next chapter. Although the findings are consistent with the previously proposed concerted model of ATP hydrolysis, the next chapter will demonstrate the superiority of a stochastic hydrolysis mechanism.

Most strikingly, analysis of the pre-steady state data of Pi accumulation demands that the portions of the cycle following hydrolysis be considered as multi-step processes. The two steps following hydrolysis are relatively slow, leading to the delay phase observed in the data. This delay phase is

Figure 2.13 – Final Catalytic Model of GroEL/GroES Concluded from Analysis of Pre-Steady State Kinetics of Inorganic Phosphate Accumulation



The arrows represent the catalytic processes: **ATP/GroES binding**, **ATP hydrolysis**, **ADP release 1** and, **ADP release 2**. The orange dotted box encloses the **symmetric cycle** and the blue dotted box encloses the states of the **asymmetric cycle**. States are numbered for convenient reference. The thickened arrows indicate that either ring can undergo the indicated process and the results would be indistinguishable, so the rate constant at this step must be double its normal value. The catalytic steps represented by the **ADP release 1** and **ADP release 2** processes, and the species represented by states 7 and 8, are unknown. The **red rings** of States 7 and 8 are to denote the uncertainty of the species being represented. Speculation of the nature of these species is made at the conclusion of the chapter.

only observed in the simulated data when additional states are included into the kinetic scheme, and only when the additional state is placed after the hydrolysis step. This is the portion of the cycle that corresponds to the release of ADP, and calls into question the assumption regarding the concerted nature of ADP release. The modified catalytic mechanism that the analysis has concluded with is presented in Figure 2.13.

The assumption of concerted ADP release is not necessarily inconsistent with the results presented in this chapter. The failure of the initial models in Figure 2.2 and Figure 2.10 to accurately simulate the data was the assumption that ADP release was a single step that followed hydrolysis. Although the analysis shows that the addition of a state in this portion of the cycle drastically improves the fit of the model, allowing the simulated data to adopt the correct concavity profile, it does not provide any indication of what species that new state represents. Concerted ADP release preceding or succeeding the additional state, as well as a multi-step ADP release process, are all consistent with the results obtained.

Though the analysis does not indicate the identity of the additional species, the notion of a slow allosteric relaxation (R → T) provides a reasonable explanation. It has been well established that the T → R allosteric transition is a rapid, concerted movement that plays a fundamental role in the positively cooperative binding of ATP. Recent studies have demonstrated that the release of ADP is accelerated by the presence of SP. These results are consistent with the idea that the R → T allosteric relaxation is an important factor in determining the rate of ADP release. SP binds to the *trans* ring of GroEL, favoring the T state, which, in turn, causes an increase in the rate of ADP release because of the decreased nucleotide affinity of the state. In the absence of SP, the slow, potentially multi-step, allosteric relaxation precedes or accompanies ADP release, drastically suppressing the overall turnover rate. These hypotheses suggest the importance of the allosteric relaxation in governing the turnover rate of the system in the presence and absence of SP, providing the system with the ability to respond to the folding catalysis demands of the cell.

2.5 REFERENCES

1. Gray, T. E. and A. R. Fersht. "Cooperativity in Atp Hydrolysis by Groel Is Increased by Groes." *FEBS Lett* 292, no. 1-2 (1991): 254-8.

2. Bochkareva, E. S., N. M. Lissin, G. C. Flynn, J. E. Rothman and A. S. Girshovich. "Positive Cooperativity in the Functioning of Molecular Chaperone Groel." *J Biol Chem* 267, no. 10 (1992): 6796-800.

3. Yifrach, O. and A. Horovitz. "Two Lines of Allosteric Communication in the Oligomeric Chaperonin Groel Are Revealed by the Single Mutation Arg196-->Ala." *J Mol Biol* 243, no. 3 (1994): 397-401.

4. Yifrach, Ofer and A. Horovitz. "Nested Cooperativity in the Atpase Activity of the Oligomeric Chaperonin Groel." *Biochemistry* 34, no. 16 (1995): 5303.

5. Langer, T., G. Pfeifer, J. Martin, W. Baumeister and F. U. Hartl. "Chaperonin-Mediated Protein Folding: Groes Binds to One End of the Groel Cylinder, Which Accommodates the Protein Substrate within Its Central Cavity." *EMBO J* 11, no. 13 (1992): 4757-65.

6. Viitanen, P. V., A. A. Gatenby and G. H. Lorimer. "Purified Chaperonin 60 (Groel) Interacts with the Nonnative States of a Multitude of Escherichia Coli Proteins." *Protein Sci* 1, no. 3 (1992): 363-9.

7. Landry, S. J. and L. M. Gierasch. "Polypeptide Interactions with Molecular Chaperones and Their Relationship to in Vivo Protein Folding." *Annu Rev Biophys Biomol Struct* 23, (1994): 645-69.

8. Fenton, W. A., Y. Kashi, K. Furtak and A. L. Horwich. "Residues in Chaperonin Groel Required for Polypeptide Binding and Release." *Nature* 371, no. 6498 (1994): 614-9.

9. Lin, Z., F. P. Schwartz and E. Eisenstein. "The Hydrophobic Nature of Groel-Substrate Binding." *J Biol Chem* 270, no. 3 (1995): 1011-4.

10. Hunt, J. F., A. J. Weaver, S. J. Landry, L. Gierasch and J. Deisenhofer. "The Crystal Structure of the Groes Co-Chaperonin at 2.8 a Resolution." *Nature* 379, no. 6560 (1996): 37-45.

11. Landry, S. J., J. Zeilstra-Ryalls, O. Fayet, C. Georgopoulos and L. M. Gierasch. "Characterization of a Functionally Important Mobile Domain of Groes." *Nature* 364, no. 6434 (1993): 255-8.

12. Rye, H. S., S. G. Burston, W. A. Fenton, J. M. Beechem, Z. Xu, P. B. Sigler and A. L. Horwich. "Distinct Actions of Cis and Trans Atp within the Double Ring of the Chaperonin Groel." *Nature* 388, no. 6644 (1997): 792-8.

13. Ranson, N. A., G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich and H. R. Saibil. "Atp-Bound States of Groel Captured by Cryo-Electron Microscopy." *Cell* 107, no. 7 (2001): 869-79.

14. Ranson, Neil A., Daniel K. Clare, George W. Farr, David Houldershaw, Arthur L. Horwich and Helen R. Saibil. "Allosteric Signaling of Atp Hydrolysis in Groel–Groes Complexes." *Nature Structural & Molecular Biology* 13, no. 2 (2006): 147-152.

15. Ma, J., P. B. Sigler, Z. Xu and M. Karplus. "A Dynamic Model for the Allosteric Mechanism of Groel." *J Mol Biol* 302, no. 2 (2000): 303-13.

16. Kipnis, Y., N. Papo, G. Haran and A. Horovitz. "Concerted Atp-Induced Allosteric Transitions in Groel Facilitate Release of Protein Substrate Domains in an All-or-None Manner." *Proceedings of the National Academy of Sciences* 104, no. 9 (2007): 3119-3124.

17. Hyeon, G. H. Lorimer and D. Thirumalai. "Dynamics of Allosteric Transitions in Groel." *PNAS* 103, no. 50 (2006): 18939.

18. Madan, D., Z. Lin and H. S. Rye. "Triggering Protein Folding within the Groel-Groes Complex." *Journal of Biological Chemistry* 283, no. 46 (2008): 32003-32013.

19. Tyagi, Navneet K., Wayne A. Fenton and Arthur L. Horwich. "Atp-Triggered Adp Release from the Asymmetric Chaperonin Groel/Groes/Adp7 Is Not the Rate-Limiting Step of the Groel/Groes Reaction Cycle." *FEBS Letters* 584, no. 5 (2010): 951-953.

20. Sameshima, T., R. Iizuka, T. Ueno and T. Funatsu. "Denatured Proteins Facilitate the Formation of the Football-Shaped Groel-(Groes)2 Complex." *Biochem J* 427, no. 2 (2010): 247-54.

21. Yang, D., X. Ye and G. H. Lorimer. "Symmetric Groel:Groes2 Complexes Are the Protein-Folding Functional Form of the Chaperonin Nanomachine." *Proc Natl Acad Sci U S A* 110, no. 46 (2013): E4298-305.

22. Ye, X. and G. H. Lorimer. "Substrate Protein Switches Groe Chaperonins from Asymmetric to Symmetric Cycling by Catalyzing Nucleotide Exchange." *Proc Natl Acad Sci U S A* 110, no. 46 (2013): E4289-97.

23. Xu, Z., A. L. Horwich and P. B. Sigler. "The Crystal Structure of the Asymmetric Groel-Groes-(Adp)7 Chaperonin Complex." *Nature* 388, no. 6644 (1997): 741-50.

24. Papo, N., Y. Kipnis, G. Haran and A. Horovitz. "Concerted Release of Substrate Domains from Groel by Atp Is Demonstrated with Fret." *J Mol Biol* 380, no. 4 (2008): 717-25.

25. Ueno, T., H. Taguchi, H. Tadakuma, M. Yoshida and T. Funatsu. "Groel Mediates Protein Folding with a Two Successive Timer Mechanism." *Mol Cell* 14, no. 4 (2004): 423-34.

26. Grason, J. P., J. S. Gresham and G. H. Lorimer. "Setting the Chaperonin Timer: A Two-Stroke, Two-Speed, Protein Machine." *Proc Natl Acad Sci U S A* 105, no. 45 (2008): 17339-44.

27. Brune, M., J. L. Hunter, J. E. Corrie and M. R. Webb. "Direct, Real-Time Measurement of Rapid Inorganic Phosphate Release Using a Novel Fluorescent Probe and Its

Application to Actomyosin Subfragment 1 Atpase." *Biochemistry* 33, no. 27 (1994): 8262-71.

28. Weiβe, A. "Global Sensitivity Analysis of Ordinary Differential Equations." Free University of Berlin, 2009.

29. Spiegel, M. *Schaum's Outline of Theory and Problems of Laplace Transforms* Schaum's Outline. New York: McGraw-Hill, 1965.

30. Lin, Zong and Hays S. Rye. "Groel-Mediated Protein Folding: Making the Impossible, Possible." *Critical Reviews in Biochemistry and Molecular Biology* 41, no. 4 (2006): 211-239.

31. Horwich, A. L., G. W. Farr and W. A. Fenton. "Groel-Groes-Mediated Protein Folding." *Chem Rev* 106, no. 5 (2006): 1917-30.

32. Horwich, Arthur L., Adrian C. Apetri and Wayne A. Fenton. "The Groel/Groes Cis Cavity as a Passive Anti-Aggregation Device." *FEBS Letters* 583, no. 16 (2009): 2654-2662.

33. Horowitz, G. H. Lorimer and Ybarra. "Groes in the Asymmetric Groel Groes Complex Exchanges Via an Associative Mechanism." *PNAS* 96, no. 6 (1999): 2682.

34. Tossavainen, T. "On the Zeros of Finite Sums of Exponential Functions." *Mathematical Society Gazette* 1, no. 33 (2006): 47-50.

Chapter 3

Analysis of Pre-Steady State Kinetics of $GroEL_7$/GroES Association

3.1 INTRODUCTION

3.1.1 Background

The ATPase activity of GroEL has been the focus of countless studies over the past 20 years. The steady state behavior has been shown to conform to an elegant model of cooperativity that incorporates the concerted principles of a positively cooperative MWC model together with the sequential principles of a negatively cooperative KNF model (1-3). However, this steady state turnover behavior is strictly a reflection of the binding of ATP and the subsequent binding coupled T → R allosteric transition, rather than the process of ATP hydrolysis.

Although early studies attributed the observed cooperativity to the hydrolysis of ATP (4), it has since been demonstrated that the binding of ATP is the cooperative process that drives the initial allosteric transition (5, 6). Much like ADP release, the treatment of the hydrolysis of ATP has been subject to the extension of the assumption of concerted domain movements (7, 8). This assumption has proliferated, in large part, due to the fact that kinetics of Pi release are well described by a single step, as was demonstrated in the previous chapter. This chapter addresses this assumption and develops a mechanistic model of ATP hydrolysis.

3.1.2 Overview

In the previous chapter, the pre-steady state kinetics of the GroEL/GroES system were monitored by performing rapid mixing experiments and measuring the release of inorganic phosphate. Analysis of the experimental data led to the conclusion that the portion of the cycle that corresponds to the release of ADP from the complex occurs in at least two slow steps. These steps are preceded by the hydrolysis of ATP to ADP and inorganic phosphate. The kinetics of the hydrolysis process were shown to be accurately described by a single step. However, this does not necessarily validate the assumption that hydrolysis is a two-state, concerted process.

This chapter will build onto the results established in the last chapter, comparing those results to the data obtained by monitoring the population of GroEL/GroES complex using a FRET pair system. These rapid mixing experiments were performed by my colleague Xiang Ye and previously published (9, 10). Much like the last chapter, the pre-steady state FRET data will be analyzed relative to a starting catalytic model: the one that the last chapter was concluded with, Figure 2.13. Adjustments to the minimal model will be made, when necessary, by pinpointing the specific area of the observed data that it fails to accurately describe, and modifying the model accordingly.

A surprising discrepancy arises between the results of this analysis and the one of the previous chapter, and the resolution to that discrepancy leads to the mechanism of ATP hydrolysis. The different hydrolysis mechanisms are then compared using theoretical single molecule simulations. The results of this chapter will then be expanded in the next chapter to include the release of GroES from the symmetric GroEL/$GroES_2$ complex.

3.2 METHODS

3.2.1 Experimental Methods

In addition to monitoring the accumulation of Pi in the system, FRET measurements can be used to determine the size of the population of $GroEL_7$/GroES complex in rapid mixing experiments. This allows for the same catalytic cycle to be monitored in two different ways. These experiments were also performed in a stop-flow apparatus. The first syringe contained apo-GroEL labeled with a donor fluorophore, $GroEL_D$, and the second syringe contained an excess of ATP and GroES labeled with an acceptor fluorophore, $GroES_A$. The two are rapidly mixed at $t = 0$. The FRET signal observed is then proportional to the number of GroES bound per GroEL. Note that this method allows for the direct observation of the population of a sum of states of the system, whereas in the last chapter the flux of a state(s) was being observed.

The observed fluorescence signal is directly proportional to the population of $GroEL_7$/GroES complex present. The output of the system can be calibrated by performing a series of experiments in the presence of BeF_x. In the presence of BeF_x and ATP, the entire population of GroEL particles forms a population of stable symmetric GroEL/$GroES_2$ complexes. This is thought to occur because BeF_x occupies the same position as the γ-phosphate of ATP after hydrolysis, effectively creating a non-hydrolysable ATP analog. In the presence of BeF_x and ADP, the population of GroEL molecules can be found in the asymmetric GroEL/GroES complex. Symmetric particles do not form under these conditions. This allows for the FRET signal to be calibrated for to the signal of one and two GroES per GroEL. All data presented below has been adjusted to this calibrated scale of moles of GroES bound per mole GroEL.

3.2.2 Model Specific Fits

As in the previous chapter, we will make extensive use of model specific fitting. See Section 2.2.5 for details.

3.2.3 Single Molecule Simulations

The model specific fits previously discussed are an example of deterministic simulations of a system. Given a kinetic scheme, a set of micro-constant values, and a set of initial conditions, the behavior of the system at all future times can be determined by solving the differential matrix equation, Appendix A.2.2. The solution corresponds to the ensemble averaged behavior of particles in the system. Each individual particle, however, will behave slightly differently, but those differences will average out and produce the bulk properties that are measured in ensemble experiments. The behavior of individual particles in the system can also be simulated. This is considered a stochastic simulation and the results of this type of simulation will vary from trial to trial. To avoid confusion, the stochastic simulations performed will be referred to as “single

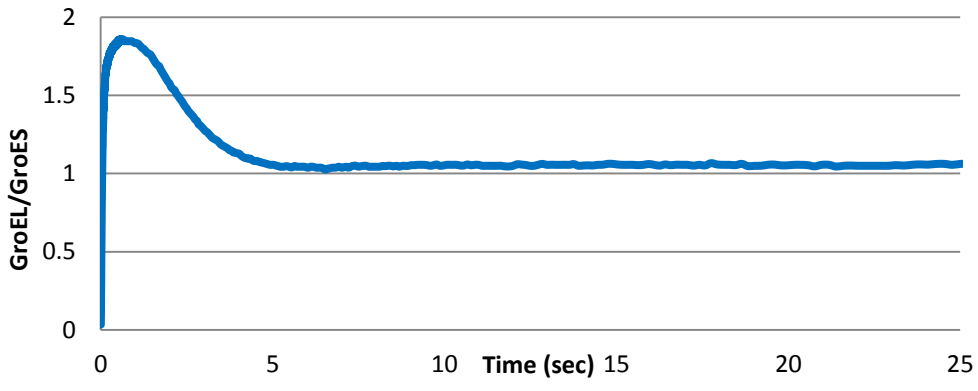
molecule" simulations. The bulk averaged results of the deterministic simulation can be obtained by averaging out a large set of single molecule simulations.

The single molecule simulations performed rely on a computer to "roll the dice" in order to randomly determine the behavior of the system. Events are simulated such that the distribution of outcomes matches what would be expected for the system when averaged over a large number of trials. For example, the lifetime of a species in a kinetic model is known to be exponentially distributed, so the species actual lifetime in a given trial is randomly selected from a set of possible times that is exponentially distributed around the proper ensemble mean. In this way, we can simulate the behavior of a single particle moving through its kinetic scheme, rolling the dice at each step to determine the length of time it spends in a particular state and, which route to take when multiple paths exist out of a single state.

3.3 ANALYSIS

An example of the data obtained from the FRET based rapid mixing experiments, performed by Xiang Ye and described above, is given in Figure 3.1. The initial state of the system is the apo state, so the data trace begins at zero. The FRET signal increases from one GroES per GroEL to

Figure 3.1 – Pre-Steady State Kinetics of $GroEL_7$/GroES Complex Formation

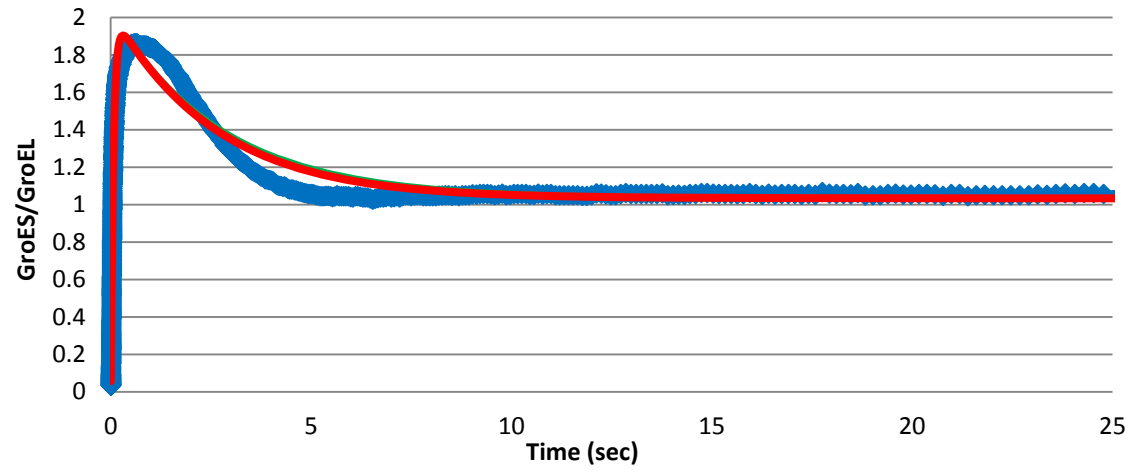


The pre-steady state kinetics of **$GroEL_7$/GroES complex formation**, obtained by Xiang Ye. GroEL, labeled with donor fluorophore, was rapidly mixed with GroES, labeled with acceptor fluorophore, and ATP. Formation of the symmetric complex (Figure 2.13, State 3) is transiently observed before the system decays to the steady-state in the asymmetric cycle.

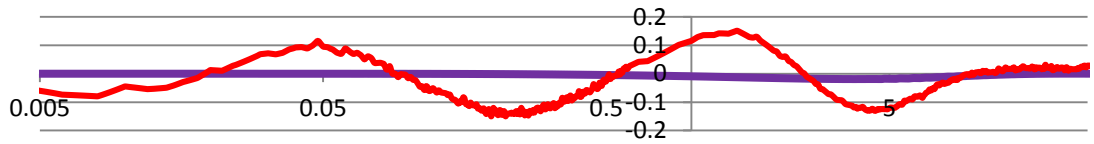
almost 2 GroES per GroEL as the system passes through the symmetric football species, State 3 of Figure 2.13. The FRET signal then decay's back to one as the system relaxes into the asymmetric cycle, where all species have only one bound GroES per GroEL molecule. Note that all species in the system, with the exception of the apo protein and the symmetric football complex, have only one GroES per GroEL.

Figure 3.2 – Results of Model Specific Fit

A



B



k_1	14	k_3	0.14
k_2	0.18	k_4	0.036
R^2	0.985		

(A) Results of the **model specific fit** of the pre-steady state **GroEL$_7$/GroES complex data**, obtained by Xiang Ye, to the model presented in Figure 2.13. Also plotted is the simulated data using the **single passage kinetic scheme** of Image 3.1 (B) The residuals of the **model specific fit** are plotted on a log-scale. Also plotted in (B) is the **difference** between the simulated **model specific fit** curve and the **single passage kinetic scheme** curve, obtained from using the fitted values of only the first two micro-constants. The two curves are nearly identical, despite the simplicity of the single passage kinetic scheme.

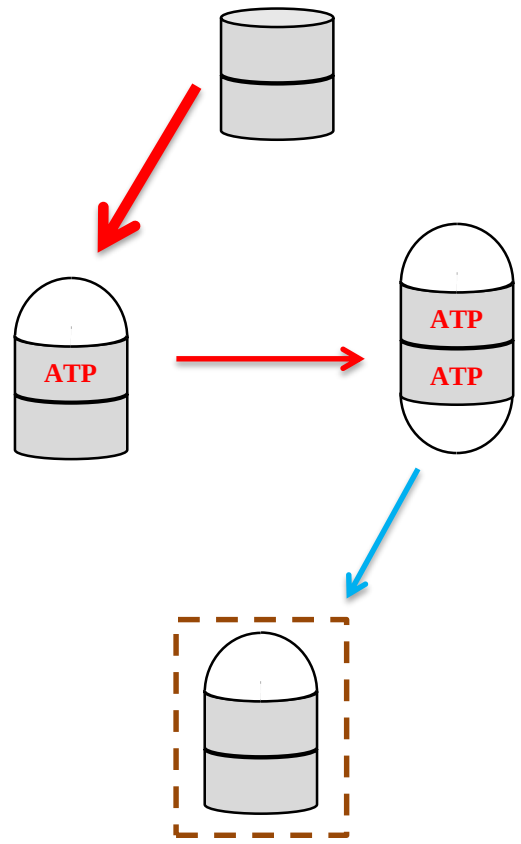
3.3.1 Model Specific Fit

A model specific fit can be performed on the data, comparing it to the kinetic scheme we concluded with in Chapter 2, Figure 2.13. Due to the complexity of the system, we must immediately resort to model specific fitting as a means of evaluating the validity of the proposed

model. The results of the model specific fit are given in Figure 3.2. With the exception of the steady state portion of the graph, the fit is very poor and the residuals are very poorly randomized.

The following analysis will identify a single step in kinetic scheme of Figure 2.13 as responsible for its inability to correctly simulate the observed data. The argument will progress by first demonstrating that, due to the observable being monitored and the experimental conditions chosen, the pre-steady state kinetics measured approximately corresponds to only a single pass through the first three states of the system. Of those three states, the problematic one can be identified by changing the experimental conditions in such a way that selectively populates that state. An increased in the observed discrepancy between theory and experiment under these conditions allow us to conclusively identify it as the part of the kinetic scheme responsible for the poor fit.

Image 3.1 – Single Passage Kinetic Scheme



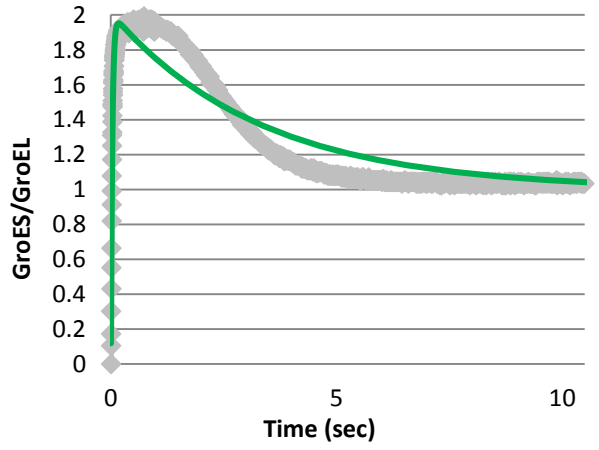
Kinetic scheme made up of a single passage through the first three states in Figure 2.12 before entering into the steady-state. The arrows correspond to the processes of **ATP/GroES binding** and **ATP hydrolysis**. The blank structure in the brown dotted box is meant to represent all the **asymmetric complexes in the steady-state**. . The thickened arrows correspond to processes that can take place on either ring, and the rates of these processes are twice the normal value.

Before changing our model to more accurately reflect the data, the problematic portion of the model must be pinpointed. First, note that the only species in the proposed kinetic scheme that contains 2 GroES per GroEL is the third state of the system in Figure 2.13. Thus, any measurement of FRET efficiency greater than 1 can be attributed to the population of the symmetric GroEL/$GroES_2$ complex, State 3,

Figure 2.13. Second, the experimental conditions used force the system almost entirely into the asymmetric cycle after its first pass. This is supported by the steady state FRET efficiency measured, which indicates that only ~5% of the steady state population is found in the symmetric state. We seek to demonstrate that the experimental data of Figure 3.2 approximately corresponds to the first passage of the system from State 1 to State 3 Figure 2.13, before entering into the asymmetric cycle where all complexes have only a single bound GroES.

The validity of this approximation can be tested by comparing the simulated data obtained from the model in Figure 2.13 to the simplified version in Image 3.1. Image 3.1 has the first three states identical to Figure 2.13, which is then connected to a single state that represents the steady-state of the system. Using only the first two fitted parameters from Figure 3.2 and the kinetic scheme of a single passage (Image 3.1), GroES occupancy data was simulated and can be found in green in Figure 3.2. Despite drastically decreasing the complexity of the system, reducing the number of states and independent parameters, the simplified kinetic scheme in Image 3.1 produces approximately the same pre-steady state kinetic trace as the one calculated using the full scheme. This demonstrates that the problematic portion of the model is localized in one or more of the first three steps in the kinetic scheme of Figure 2.13.

Figure 3.3 – Model Specific Fit Under Conditions of Increased [GroES]



Pre-steady state **GroEL/GroES complex data**, obtained by Xiang Ye, collected under conditions of increased [GroES]. A **model specific fit** was performed using the **single passage kinetic scheme** in Image 3.1. The hydrolysis coupled decay of the symmetric particle is inadequately described by the model.

The only two processes in the first three steps of the system are the two binding events and the hydrolysis event. In order to demonstrate that the binding steps are not the cause of the discrepancy, an additional experiment was performed by my colleague Xiang Ye, under conditions of increased

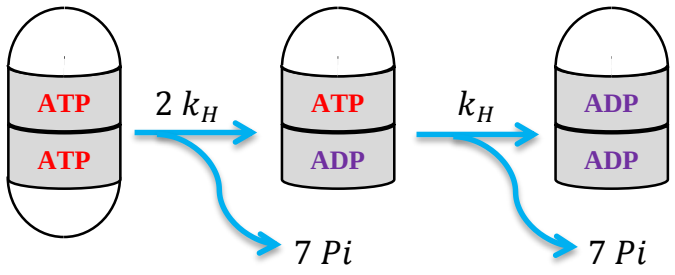
[GroES]. If the approximation that combines ATP and ES binding into a single step is the issue, the fit should improve under these conditions. Figure 3.3 shows the results of the model specific fit to the system under this new set of experimental conditions. The fit is dramatically worse than the fit presented in Figure 3.2, suggesting that the assumption that ATP and GroES binding occurs in a single step is not the problem.

Under the conditions of high [GroES], population of the symmetric state occurs almost instantaneously and completely. Looking at Figure 3.3, it is obvious that the model is inadequately suited to describe the decay of the population of symmetric particles. Inspection of the fit shows that when approximating the hydrolysis step as a single exponential decay, the transition begins too quickly and then finishes more slowly than the true data. The true population of symmetric GroEL/$GroES_2$ complexes exists for longer, and then decays more abruptly than a single exponential model will allow. The experimental data shows that the hydrolysis process is inadequately described by as a one-step process, and instead requires multiple states.

3.3.2 Symmetric Complex Population vs Hydrolysis Events

Another way of represent the inadequacy of the concerted, single-step hydrolysis assumption in characterizing the decay of the GroEL/$GroES_2$ population is to plot the population of symmetric particles against the average number of hydrolysis events. The plot is independent of the value of the hydrolysis micro-constant chosen, and is thus a good way to look at the hydrolysis step without having to worry about the accuracy of the micro-constant value. In the concerted hydrolysis model, hydrolysis within a ring was proposed to occur as a single step. This means that the hydrolysis of all 14 ATP molecules can be expressed as the three state kinetic scheme in Image 3.2. The analytical solution to the kinetic scheme can be easily solved using the inverse

Image 3.2 – Concerted Hydrolysis



A three state linear kinetic model of hydrolysis concerted within a ring. The system can be solved analytically and the population of each species is plotted against number of Pi's released in Figure 3.4.

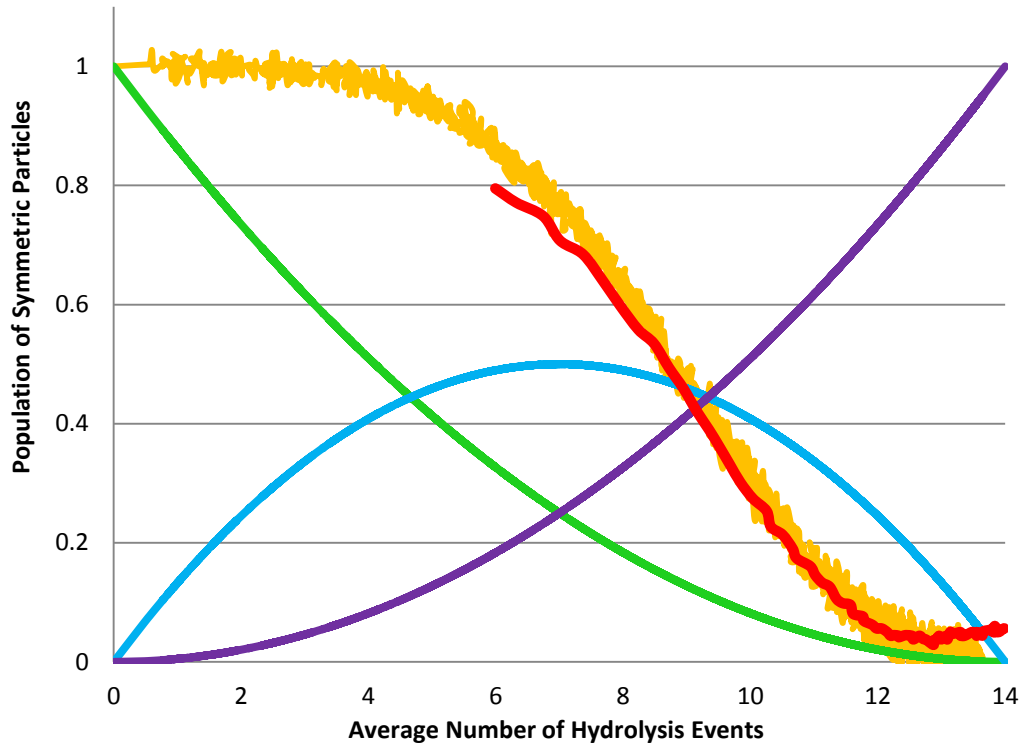
Laplace transform method outlined in the Appendix A.2.4. Of the three states in the system, only $X_1(t)$ corresponds to the population of symmetric particles and it can be plotted against the sum of the flux functions $Y(t) = \frac{1}{2}[Y_1(t) + Y_2(t)]$ for a collection of $t \in [0, \infty)$.

A plot of the population of symmetric particles against number of hydrolysis events was achieved experimentally by collecting pre-steady state traces of both FRET efficiency and Pi release as a function of time, and plotting the measurements taken at particular time points against one another, Figure 3.4. As is demonstrated by Figure 3.4, the start of the decay in the population of symmetric GroEL/$GroES_2$ complexes simulated from the concerted kinetic scheme in Image 3.2 occurs simultaneously with hydrolysis. This is not supported by the experimental data, which indicates that the symmetric complex survives multiple rounds of hydrolysis before beginning to decay. This indicates that GroEL/$GroES_2$/ATP_{14} is not the only symmetric species present in the catalytic cycle. The population of symmetric complexes persists throughout a large number of hydrolysis events, which implies that State 3 of Figure 2.13 is representative of a set of multiple distinct nucleotide-bound symmetric species.

3.3.3 Equating Nucleotide Symmetry and GroEL/$GroES_2$ Complex Symmetry

Thus far, we have assumed that GroES release from the symmetric $(ADP)_7/(ATP)_7$ is extremely rapid. This allows us to equate the loss of nucleotide symmetry, i.e. going from $(ATP)_7/(ATP)_7$ to $(ADP)_7/(ATP)_7$, as equivalent to the loss of complex symmetry, i.e. going from GroEL/$GroES_2$ to GroEL/GroES, Image 3.3.A. An alternative explanation to the above argument against concerted

Figure 3.4 – Comparison of Population of Concerted Hydrolysis Species with Increasing Average Number of Hydrolysis Events and Experimental Data

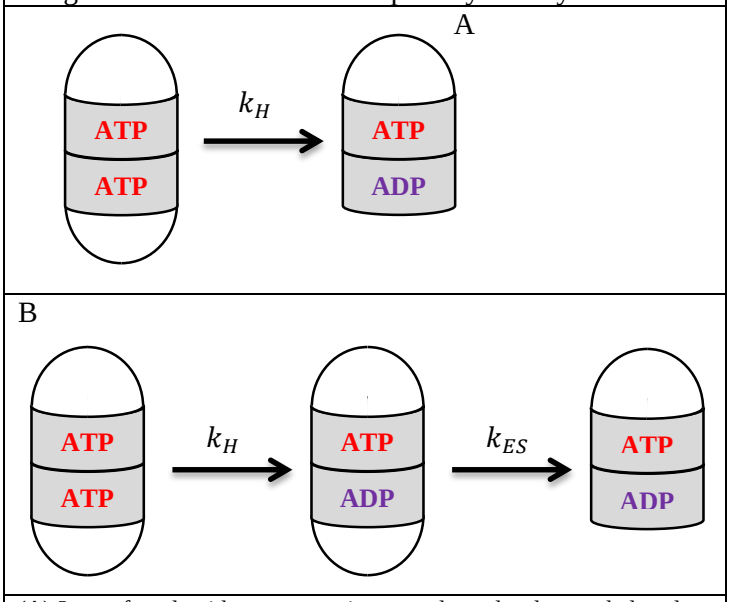


This plot is a comparison of the population of species in the concerted hydrolysis model, Image 3.2, and the experimental data, collected by Xiang Ye, of the decay of the symmetric GroEL/GroES$_2$ complex. There are three species in the concerted hydrolysis model: the **symmetric GroEL/GroES$_2$ complex**, **the asymmetric GroEL/GroES/ATP$_7$/ADP$_7$** and the **asymmetric resting state complex GroEL/GroES/ADP$_{14}$**. The experimentally observed decrease in the GroEL/GroES$_2$ complex is plotted against Pi accumulation for **wild-type GroEL** and the **D398A mutant**. The experimentally observed decay of the symmetric particle has a large lag, indicating its survival throughout numerous rounds of hydrolysis. In contrast, the concerted model predicts that the **symmetric complex** decays after a single hydrolysis event, indicated by the immediate decrease in the symmetric population and increase in the asymmetric population.

hydrolysis is that hydrolysis is indeed concerted, but followed by one or more species that precede GroES release from the symmetric complex, Image 3.3.B. The fault in the model in Figure 2.13 would then not be the fact that hydrolysis is represented as concerted, but in the equating nucleotide and complex symmetry.

The slow-hydrolyzing GroEL mutant, D398A, can be used to experimentally demonstrate the validity of equating nucleotide symmetry and complex symmetry. If it were the case that hydrolysis is concerted and GroES release from the symmetric complex occurs in multiple steps then, in the limit of slow hydrolysis, these other processes would proceed relatively quickly,

Image 3.3 – Nucleotide vs Complex Symmetry



(A) Loss of nucleotide symmetry is equated, or closely coupled, to loss of complex symmetry with respect to bound GroES. (B) Loss of complex symmetry is taken as a separate step, distinct from loss of nucleotide symmetry.

allowing us to equate nucleotide symmetry and complex symmetry. This corresponds to the situation where $k_H \ll k_{ES}$ of Image 3.3.B, which, by the argument laid out in Appendix A.2.3, allows us to approximate its kinetic scheme as equivalent to that of the model in Image 3.3.A. The D398A mutant of GroEL hydrolyzes ATP at about 0.1% of the wild-type rate, allowing us to establish the link between nucleotide and complex symmetry.

Figure 3.4 contains experimental plots of the decay of the symmetric complex of both wild-type GroEL/GroES$_2$ and D398A/GroES$_2$, acquired by Xiang Ye. Hydrolysis in D398A is many orders of magnitude slower than binding, allowing a full population of symmetric D398A/GroES$_2$ complexes to form and the first hydrolysis events to be resolved. This is in contrast to the wild-type case, where binding competes with the first few hydrolysis steps. Nonetheless, Figure 3.4 demonstrates that the decay of the symmetric wild-type and D398A complexes precede through the same hydrolysis steps, despite the large variation in the hydrolysis rates. If equating nucleotide and complex symmetry was the source of error in our model specific fits, the decay of the symmetric wild-type complex would be shifted to the right of the D398A curve, because its survival throughout multiple hydrolysis events is a consequence of the slow decay of the complex. Furthermore, if hydrolysis was concerted, the D398A curve would trace the simulated data for the concerted system, because the extremely slow rate of hydrolysis would dominate the

system's behavior. Neither of these predictions are consistent with the observed data, validating the equation of nucleotide and complex symmetry.

3.3.4 Summary

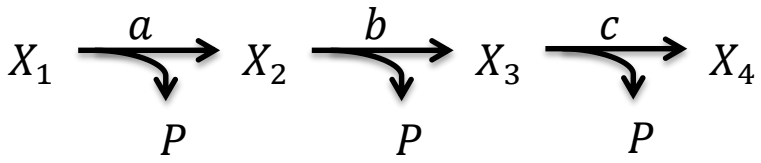
A kinetic model was obtained in Chapter 2 by rigorous analysis of the pre-steady state kinetics of inorganic phosphate release. The same model was then attempted to be applied the pre-steady state kinetics of $GroEL_7$/GroES complex formation, where it proceeded to fail miserably. The decay of the symmetric GroEL/$GroES_2$ complex was identified as the step in the cycle that the model was unequipped to describe. Analysis of the decay process in the space of hydrolysis events demonstrated that the model lacked a sufficient number of symmetric species to accurately describe the process. The model assumed the hydrolysis of ATP in the symmetric complex was concerted and thus could be described by a single symmetric species, GroEL/GroES/ATP_{14}. Finally, a comparison of the decay of the symmetric wild-type GroEL/$GroES_2$ and D398A/$GroES_2$ allowed us to conclusively tie the decay of the symmetric complex to ATP hydrolysis, effectively equating the events. Thus, we established that the hydrolysis of ATP was insufficiently described as a single-step in the model derived from Figure 2.13.

3.4 DEVELOPMENT OF HYDROLYSIS MECHANISM

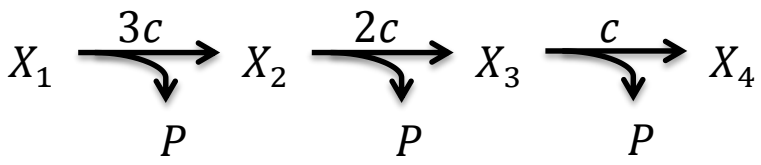
In monitoring the system using two different measures, an interesting contradiction has arisen that can be directly attributed to the hydrolysis step in the kinetic scheme. If measuring the release of Pi, hydrolysis is well described as a two state, single-step process. The sensitivity analysis of the system, Figure 2.9, demonstrated that the burst kinetic-nature of Pi accumulation effectively isolates the hydrolysis step relative to the others in the fit, leading to increased confidence in the relationship between the data and the model in this particular portion in the cycle. That is, out of all the steps in the cycle, the hydrolysis step is the one modeled with the most accuracy. However, by monitoring formation and decay of the GroEL/$GroES_2$ states of the system, it was

Image 3.4 – Comparison of Linear Kinetic Schemes

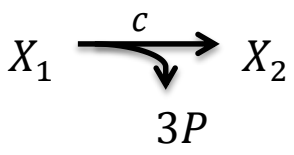
A



B



C



demonstrated that multiple species are needed to accurately describe the hydrolysis process. This discrepancy suggests that the hydrolysis of ATP is a multi-step process whose flux function reduces to a single exponential and is thus well described as a single-step. It turns out that this condition has insightful mechanistic implications for the hydrolysis process.

3.4.1 Linear Kinetic Schemes

In order to determine the relationship between micro-constants that would generate observed discrepancy, the model kinetic schemes in Image 3.4.A will be examined. This is a simple scheme with four states connected in parallel by three processes that are accompanied by the release of product. The goal is to find the relationship between the micro-constants in Image 3.4.A that would reduce the product accumulation function, the cumulative flux function of the states, to a single exponential. The kinetic scheme in Image 3.4.A was solved analytically using the inverse Laplace transform method (A.2.4) and its cumulative flux function is given as Eq 3.1. As expected, the cumulative flux function for the system contains the sum of three exponential terms. In order for the cumulative flux function to reduce to a single exponential, the coefficients of two of its exponential terms must be zero. Furthermore, the denominator of each term contains the difference between two of the system's micro-constants: equating any two micro-constants

would send these terms to infinity, and is forbidden (the case where two of the micro-constants are equal is not presented but the solution can be obtained by a different method, and it does not have the desired single exponential form).

$$Y_{1+2+3}(t) = 3 + \frac{(-3bc + 2ab - a^2 + ac)e^{-at}}{(a-b)(a-c)} + \frac{a(a-c)(2c-b)e^{-bt}}{(b-c)(a-b)} - \frac{abe^{-ct}}{(a-c)(b-c)} \quad \text{Eq 3.1}$$

First, note the coefficient out front of e^{-ct} is never zero. The coefficient of the e^{-bt} term is zero when either $a = c$ or $b = 2c$. However, $a = c$ is forbidden, implying $b = 2c$. Plugging this result back into Eq 3.1 and simplifying yields Eq 3.2. The coefficient out front of e^{-at} will be zero only when $a = 3c$. Plugging this result into Eq 3.2, the cumulative flux equation reduces to its final form, Eq 3.3.

$$Y_{1+2+3}(t) = 3 + \frac{(a-3c)e^{-at}}{(a-c)} - \frac{2ace^{-ct}}{(a-c)c} \quad \text{Eq 3.2}$$

$$Y_{1+2+3}(t) = 3 - 3\,e^{-ct} \quad \text{Eq 3.3}$$

Thus, if the relationship between the micro-constants is $\frac{a}{3} = \frac{b}{2} = c$, then the cumulative flux function of the system is reduced to a single exponential. Also, just as we had hoped, Eq 3.3 is equivalent to the flux function obtained from the system pictured in Image 3.4.C, which corresponds to a kinetic scheme where the release of 3 product molecules occurs in a concerted, single-step.

The relationship between the micro-constants obtained above, and presented in Image 3.4.B, implies a special relationship between the three product release processes. The micro-constant relationship is what one would expect if the three processes were stochastic: equivalent events that occur independently and with equal probability. Thus, in the case of linear systems, the flux function of a concerted process is equivalent to the sum of flux functions of a series of stochastic processes.

3.4.2 Cyclic Kinetic Schemes

In the case of non-linear systems, such as those that describe enzymatic catalysis, the flux function of a stochastic process is no longer exactly equivalent to that of a concerted process. Nonetheless, the equivalence does still hold exactly for the first pass through the cycle. The proof for this statement can be found in the Appendix 3.1. For most systems, the first pass is the only major contributor because after the first pass, the system has fallen almost entirely into the steady state, where the contribution of the individual exponential terms are no longer significant.

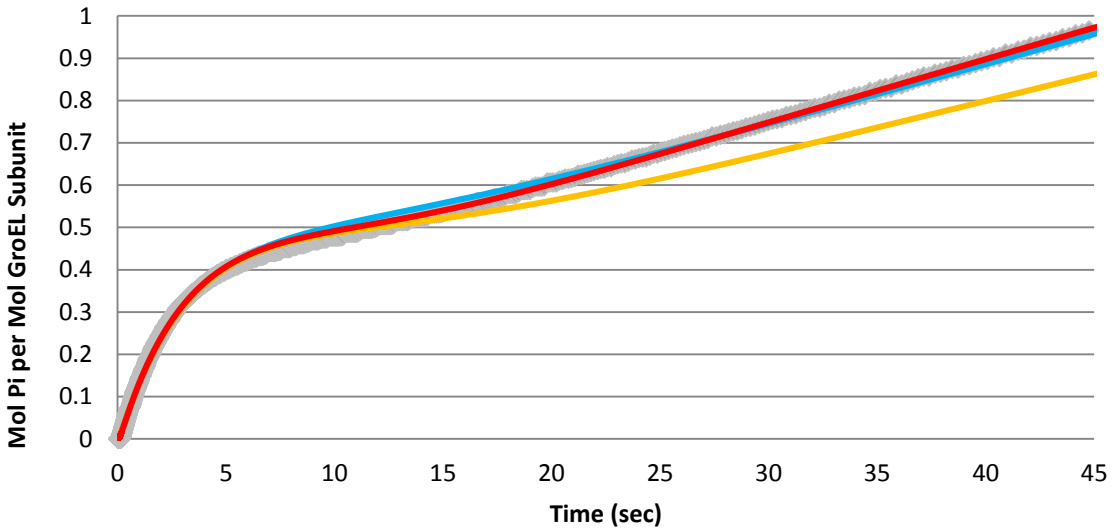
The assumption can be evaluated within the context of the system in question in order to ensure that a change to our catalytic model does not change what was established in the previous chapter. The focus of this analysis will be limited to the asymmetric cycle because expanding the hydrolysis process from a concerted to a stochastic one requires seven species per ring. Applying this assumption to the full catalytic model of Figure 2.13 introduces a large degree of complexity, which will be the subject of Chapter 4.

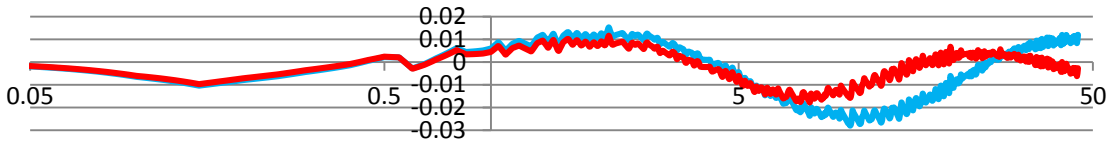
First, to demonstrate the major differences in the two models, the same micro-constants obtained from the model specific fit of the concerted kinetic scheme in Figure 2.7 are simply plugged into the stochastic kinetic scheme. This demonstrates that the fit to the beginning portion of the curve, through the burst phase, remains the same, Figure 3.5. The difference is in the steady state portion of the graph. It can be shown that, for a cyclic system of N states, such as the kinetic schemes used to describe the behavior of the asymmetric cycle, the steady state rate of product release, S, is related to the set of micro-constants, $\{k_i\}$ by Eq 3.4.

$$\frac{1}{S} = \sum_{i=1}^{N} \frac{1}{k_i} \qquad \text{Eq 3.4}$$

The addition of multiple new micro-constants will have a significant impact on the steady-state portion of the simulated data. However, the effect may be countered by changes in the values of

Figure 3.5 – Using the Stochastic ATP Hydrolysis Model for the Pre-Steady State Kinetics of Inorganic Phosphate Accumulation





Original Fit to Concerted Hydrolysis Model		New Fit to Stochastic Hydrolysis Model	
k_1 (sec^{-1})	11.1	k_1 (sec^{-1})	9.7
k_2 (sec^{-1})	0.33	k_2 (sec^{-1})	0.34
k_3 (sec^{-1})	0.063	k_3 (sec^{-1})	0.051
k_4 (sec^{-1})	0.063	k_4 (sec^{-1})	0.16
R^2	0.9983	R^2	0.9993

Model specific fit of the **Pi accumulation data**, obtained under asymmetric cycle conditions by Xiang Ye. Previous fit performed using **concerted hydrolysis model** performed in Figure 2.7. The **orange curve** was generated by inserting the micro-constant values obtained in the concerted model fitting, Figure 2.7, into the stochastic model. The orange curve is shown to demonstrate that the model change **does not affect the burst phase of the fit**. The observed deviation in the steady state behavior of the curve is predictable given the addition of several new steps in the cycle. A new **model specific fit** using **the stochastic model**, restores the goodness of fit by increasing the value of one of the smaller micro-constants. The value of the hydrolysis micro-constant, k_2, does not significantly change during the process.

the smaller rate constants to reflect the new steady state rate. The sensitivity analysis in Chapter 2 demonstrated that the smallest valued micro-constants have the most significant impact on the steady state behavior of the system. A model specific fit was be performed using the stochastic kinetic scheme to obtain a new fitted value for these micro-constants. Note that the hydrolysis micro-constant is not significantly altered in the new fit. The fitted value of the forth micro-

constant in the stochastic model is increased to account for the steady state turnover rate, leaving the goodness of the fit, measured by the R^2 value, effectively unchanged.

3.5 THEORETICAL COMPARISON OF HYDROLYSIS MECHANISMS

3.5.1 Lifetime of the Full Ring

Our analysis of the pre-steady state kinetics of GroEL has led us to the surprising result that, contrary to popular belief, the hydrolysis of ATP is likely stochastic and not concerted. The pre-steady state kinetics of the two mechanisms are nearly indistinguishable with respect to the release of inorganic phosphate but results differ drastically with respect to the release of GroES from the symmetric GroEL/$GroES_2$ complex, which is dependent on the distribution of nucleotide states.

In general, the mean residence time, τ_i, of a state can be calculated using Eq 3.5-6 where $\hat{X}_i(t)$ is the normalized form of the state function $X_i(t)$.

$$\hat{X}_i(t) = \frac{X_i(t)}{\int_0^\infty X_i(t)\,dt} \qquad \text{Eq 3.5}$$

$$\tau_i = \int_0^\infty t\,\hat{X}_i(t)\,dt \qquad \text{Eq 3.6}$$

Consider the two kinetic schemes presented in Image 3.5. The first scheme, Image 3.5.A, corresponds to a system where the processes linking the states are stochastic. The second, Image 3.5.B, is the analogous system where the process is a two-state concerted process. The relationship between the flux functions of the two systems has already been demonstrated to be equivalent. In addition, the analytical solution of each scheme can also be used to calculate the average amount of time required to hydrolyze all 7 ATP molecules: the average lifetime of a full ring. It will be shown in the following sections that this quantity is related to the lifetime of the symmetric complex under certain conditions.

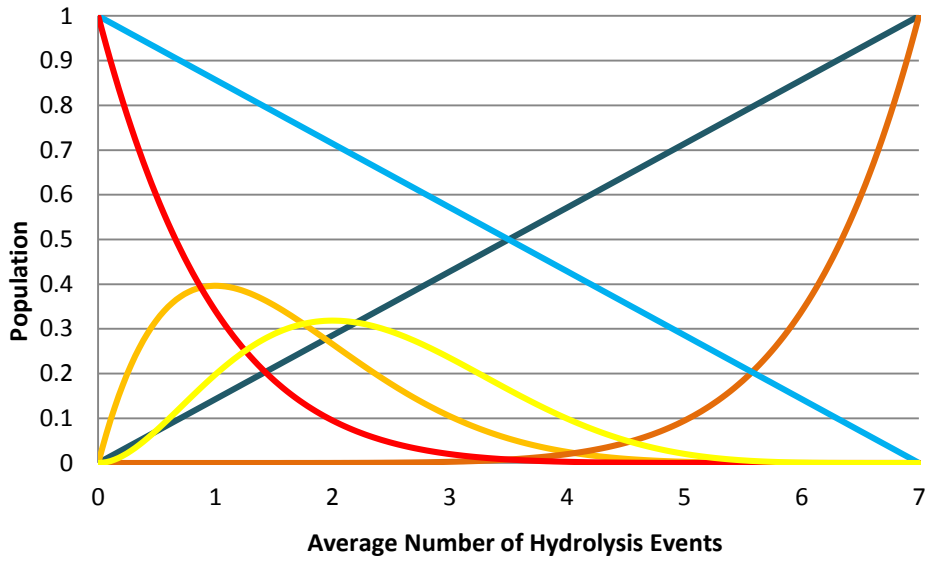
Image 3.5 – Kinetic Schemes for a Stochastic and Concerted System
A $X_1 \xrightarrow{7\,k} X_2 \xrightarrow{6\,k} X_3 \xrightarrow{5\,k} X_4 \xrightarrow{4\,k} X_5 \xrightarrow{3\,k} X_6 \xrightarrow{2\,k} X_7 \xrightarrow{k} X_8$
B $Z_1 \xrightarrow{k} Z_2$

For the concerted scheme, Image 3.5.B, the population of full rings is equal to the population of $Z_1(t)$. The average lifetime of a full ring is then simply the mean residence time of the system in state $Z_1(t)$. Using the solution to the kinetic scheme obtained by the inverse Laplace transform method, and plugging it into Eq 3.6, we obtain $\tau_{Con} = \frac{1}{k_H}$.

In the stochastic system, the process of hydrolyzing a full ring occurs over 7 states. Then, the sum of the states $X(t) = \sum_{i=1}^{7} X_i(t)$ would correspond to the population of particles in the stochastic system with nucleotide bound. However, the lifetime of a full ring would be the average amount of time required to reach the state X_8 (corresponding to the ADP_7 species), which is just the mean residence time of $X_7(t)$. Again, using the solution to the kinetic scheme obtained by the inverse Laplace transform method, and plugging it into Eq 3.6, we obtain $\tau_{Sto} = \sum_{i=1}^{7} \frac{1}{i*k_H} = \frac{363}{140}\frac{1}{k_H} \approx 2.59\ \frac{1}{k_H}$. Thus, for a given hydrolysis micro-constant, k_H , the average length of time required reach the fully hydrolyzed $(ADP)_7$ state is about 2.59 times longer for a stochastic system than a concerted system. Note: $\frac{363}{140} = \frac{1}{7} + \frac{1}{6} + \frac{1}{5} + \frac{1}{4} + \frac{1}{3} + \frac{1}{2} + 1$.

The above result that the lifetime of the full ring is, on average, shorter in the case of a concerted system compared to a stochastic system, may appear to contradict to the notion that the rate of hydrolysis of the two systems are identical. However, consider the curves in Figure 3.6 comparing the distribution of states of the concerted and stochastic kinetic schemes in Image 3.5. When an average of 1/7 ATPs have been hydrolyzed, 1/7 complexes in the concerted system have reached the $(ADP)_7$ state, Z_2, while the remaining 6/7 are still in the $(ATP)_7$ state, Z_1. In contrast,

Figure 3.6 – Comparison of the Distribution of States in a Concerted verse Stochastic Kinetic Scheme



Partial distribution of the states of the two kinetic schemes pictured in Image 3.5. The two blue curves are the calculated distributions of the concerted states Z_1 and Z_2. The theoretical distribution of the first three states of the stochastic scheme, X_1, X_2 and, X_3 are given. In addition, the last state in the stochastic scheme, X_8, is given for comparison. The figure is meant to illustrate how, despite having identical product outputs, the state distributions of the kinetic schemes are drastically different.

when 1/7 of the ATPs have been hydrolyzed in the stochastic system, only a very very small fraction of complexes, $< 10^{-5}$, have reached the state $(ADP)_7$. The vast majority, > 90%, of the population of the system can be found between states X_1, X_2 or, X_3. Despite the fact that the accumulation of inorganic phosphate is the same in both systems, the accumulation of the $(ADP)_7$ state is drastically different, occurring over narrow range of hydrolysis events in the stochastic system.

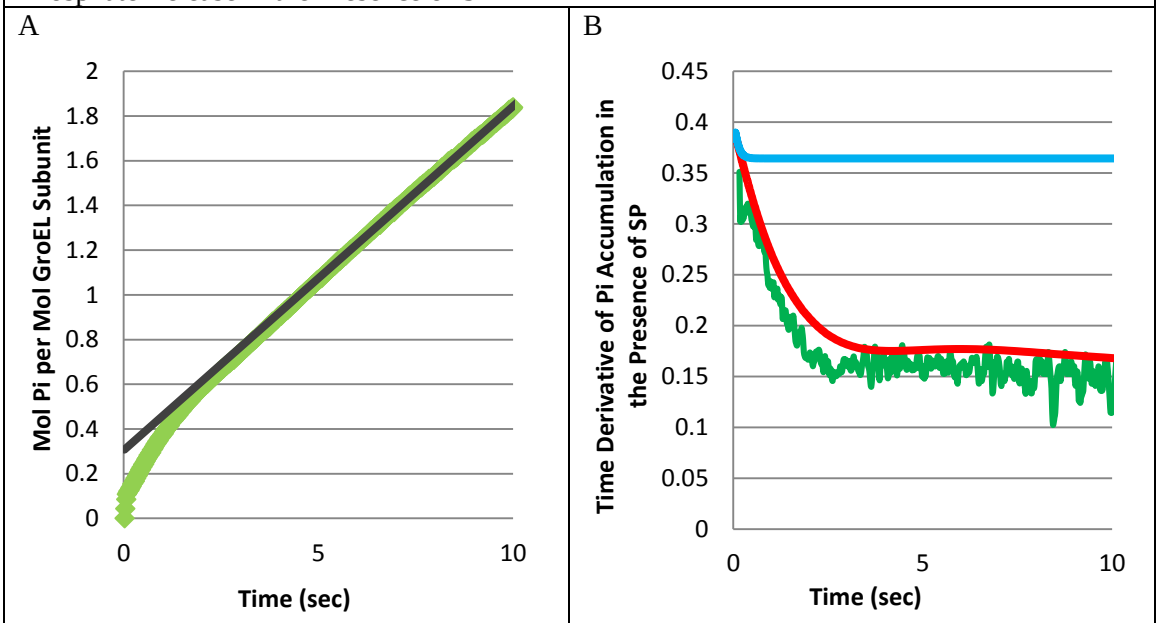
3.5.2 Hydrolysis in the Presence of SP

The calculations of the lifetime of a full ring above lead to an additional piece of evidence in support of the stochastic hydrolysis mechanism in the GroEL catalytic cycle. This can be found by analyzing the steady state and pre-steady state kinetics of inorganic phosphate accumulation in the presence of SP. Our model specific fits of the various kinetic schemes laid out in Chapter 2 led to sets of fitted micro-constants. The values of the hydrolysis micro-constant obtained were in relatively good agreement between the various fits due to the unique role of the hydrolysis micro-

constant in governing the pre-steady state burst kinetics, demonstrated by the confidence analysis in Figure 2.9. The concerted and stochastic models of hydrolysis lead to different relationships between the micro-constants and the steady-state rate of hydrolysis. However, under conditions where ADP release is the rate limiting step (absence of SP), this difference does not allow us to distinguish between the models: both can be fit to the experimental data reasonably well because the discrepancy is accounted for by one of the other micro-constants.

In contrast to the previous systems analyzed, in the presence of SP the hydrolysis step(s) are rate limiting, and the hydrolysis micro-constant becomes the dominant term in the relationship between the micro-constants and the steady-state rate turnover rate, Eq 3.4. In the concerted model, this term takes on the value k_H whereas in the stochastic model, the value is ~ $\frac{k_H}{2.59}$. Since

Figure 3.7 – Stochastic Hydrolysis is Supported by the Pre-Steady State Kinetics of Inorganic Phosphate Release in the Presence of SP



In the presence of SP, hydrolysis is the rate-limiting step in the catalytic cycle of GroEL/GroES and has increased influence over the steady state behavior of the system. (A) a **best fit line** is fit to the steady state portion of the **Pi accumulation data measured in the presence of SP**, obtained by Xiang Ye. The slope of the line corresponds to 0.386 times the hydrolysis micro-constant fitted for the system ($k_H = 0.4$), as predicted by the stochastic hydrolysis model. (B) Data was simulated using a **concerted** and **stochastic** cyclic model and the slopes are plotted along with the **numerically calculated slope of the experimental data set**. The **concerted** system enters into the steady state almost immediately whereas the **stochastic** system slowly decays, from the value k_H to 0.386 k_H, as observed in the **experimental data set**.

the value of k_H has been determined with a relatively high degree of confidence, it can be compared the steady-state rate of hydrolysis in the presence of SP. The relevant data, collected by Xiang Ye, is presented in Figure 3.7.A.

The average of the two fitted values for the hydrolysis micro-constant at 100mM K^+ is ~0.4 s^{-1}. The slope of the steady-state portion of the Pi accumulation data obtained at 100mM K^+ in the presence of SP is 0.154 s^{-1}, which is ~ $\frac{1}{2.59}$ times the value of the hydrolysis micro-constant. These results suggest that GroEL hydrolyzes ATP stochastically.

Further support for the stochastic mechanism of hydrolysis can be found by analyzing the slope of the experimental pre-steady state data compared to the curves simulated using a concerted and stochastic model of hydrolysis. The concerted model enters into the steady-state almost immediately, and the value of the slope stabilizes at k_H. For the stochastic model of hydrolysis, there is a small decrease in the slope during the pre-steady state portion of curve, as the value decreases from k_H to $\frac{140}{363}\ k_H$. This behavior is seen in both the experimental data and the curve simulated using a stochastic model of hydrolysis, but not in the concerted hydrolysis model, Figure 3.7.B.

3.5.3 Steady State Population Distributions

The rate of hydrolysis in the GroEL/GroES system is assumed to be zeroth-order and thus insensitive to the concentration of substrates. In contrast, the value of the binding micro-constant is first-order, and thus directly proportional to the concentration of ATP. Likewise, the value of the release micro-constant has been demonstrated to increase over 100-fold in the presence of SP (10). The GroEL/GroES system is expected to respond to changes in these conditions appropriately: becoming folding-active in response to an increase in aggregated/ misfolded proteins and conserving energy (i.e. not hydrolyzing ATP) when demand is low. The following

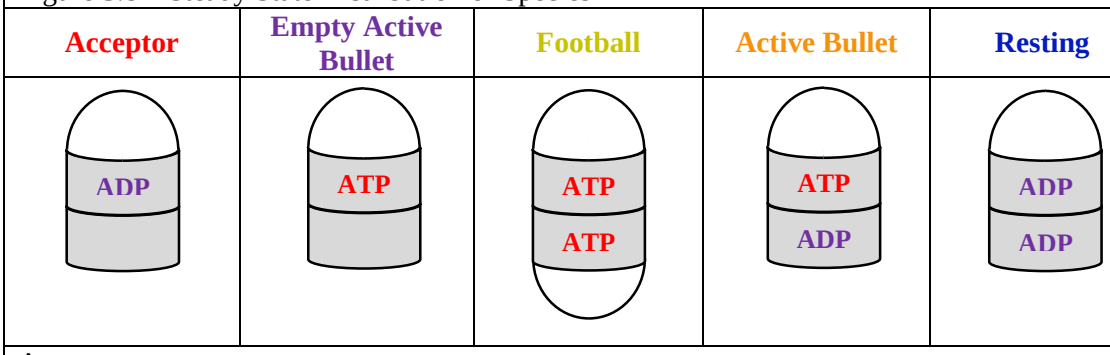
analysis compares the two hydrolysis mechanisms based on their sensitivity to the value of the micro-constants for binding and release.

In order to determine the sensitivity of the system to changes in the binding and release micro-constants, the steady-state populations of five different, catalytically relevant species were calculated for each mechanism. The species of interest in the analysis are shown in Figure 3.8. The steady-state population of each species can be calculated by using the rate matrix for the system and solving for $\frac{d\vec{X}}{dt} = 0$. The solution to the matrix equation gives the steady-state population of each species as a function of the system's micro-constants.

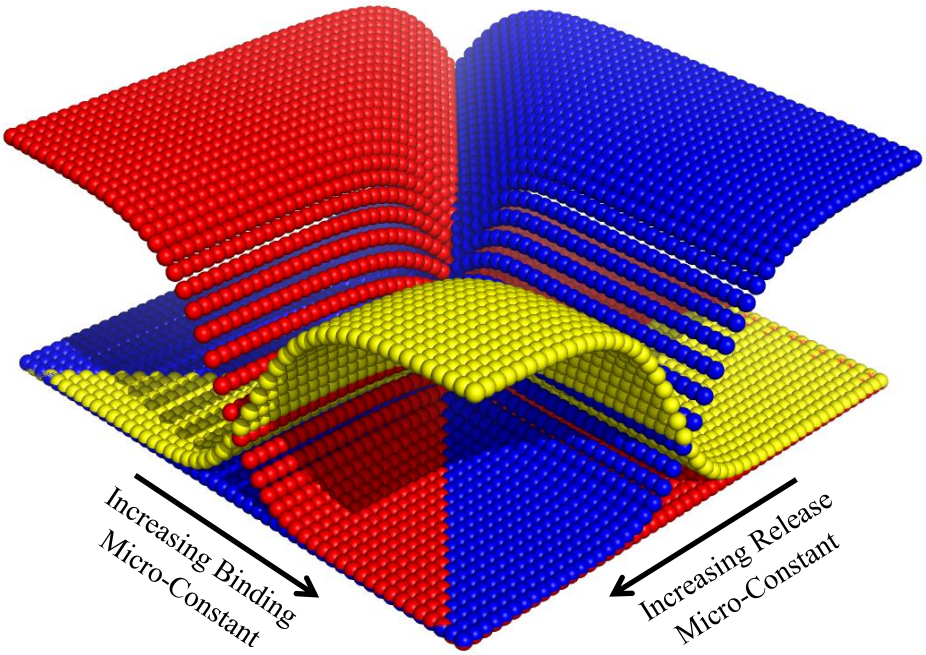
Despite the findings of the previous chapter, indicating how important it is to consider ADP release as a two-step process, the two rate constants are combined in the present analysis to limit the number of independent variables. The sum of the reciprocals of the two release micro-constants is equal to the inverse of the release parameter used the analysis, however, for simplicity, it is inaccurately referred to as the "release micro-constant". The values of the binding and release micro-constant were varied over 6 orders of magnitude, equally spaced in log-space and the populations of each of the species in Figure 3.8 were calculated. The hydrolysis micro-constant for the concerted system was set to $H_{Con} = 1$ and for the stochastic system, $H_{Sto} = {}^{363}/_{140}$. The difference in the value of the hydrolysis micro-constant used was done in order to equate the lifetime of the full ring for the two mechanisms. The results of the analysis are given in Figure 3.8.

For both hydrolysis mechanisms, the distribution of all species is, surprisingly, identical. Footballs are the dominant species at larges values of the binding and release micro-constants, corresponding to conditions of high [ATP] and high [SP], respectively. When the release micro-constant is low (absence of SP) and the binding micro-constant high (high [ATP]), the system is

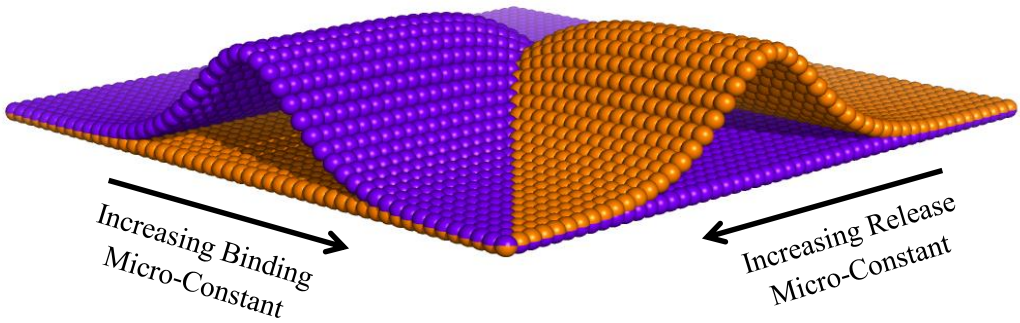
Figure 3.8 – Steady State Distribution of Species



A

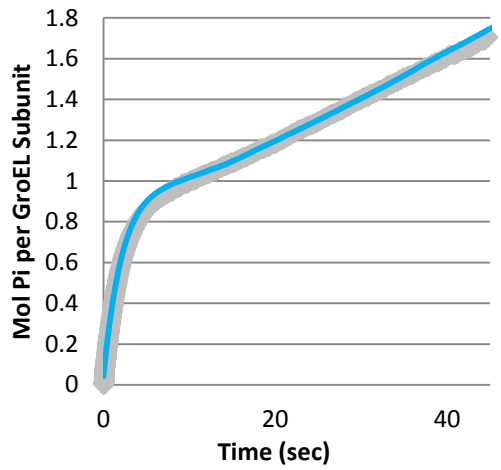


B



The steady state population of each of the five species presented at the top of the figure is given as a function of the ATP binding and ADP release micro-constant values. The values of B and R used are presented on a log scale and vary over six orders of magnitude. (A) The populations of the dominant species observed at each experimental extreme. (B) Folding active asymmetric species, which act as transition species during activation by either SP or ATP.

Figure 3.9 – Reproducing the Pre-Steady State Ensemble Data with Single Molecule Simulations



The **Pi accumulation data**, obtained by Xiang Ye, starting with apo-GroEL as the initial state, was compared to the averaged results of 1000 **single molecule simulations**. This was done in order to verify the accuracy of the single molecule simulation method used. The two curves appear to be in relatively good agreement, including the presence of the characteristic delay phase observed in the data.

pushed towards the resting state where the *trans* ring is still occupied by ADP. At low values for the binding micro-constant (low [ATP]) and large values for the release micro-constant (high [SP]), the system is in the acceptor state, where the *trans* ring is void of nucleotide. In both cases, the predominant species is one of the two asymmetric complexes. As either B or R are increased, the system passes through the asymmetric "active bullet" states, where the ATP timer on the *cis* ring is active, while the *trans* ring awaits the next round of either SP or ATP binding. Despite the differences in the hydrolysis mechanism, the steady-state distribution of species is identical, indicating that neither mechanism displays more or less sensitivity to a changes in the other micro-constants of the system.

3.5.4 Single Molecule Simulations

The first set of single molecule simulations were performed in an effort to simulate the Pi accumulation data obtained from experiment by Xiang Ye, averaging over a large set of single molecule trajectories. This was done as a check to ensure that the model used for the single molecule simulations matched the model used for the deterministic simulations. These single molecule simulations were performed using the fitted micro-constant values obtained from the model specific fits to the Pi accumulation data starting with apo-GroEL (Figure 2.10.B) and

Table 3.1 – Micro-constant Values Used in Single Molecule Simulations				
	Concerted Model		Stochastic Model	
	(-) SP	(+) SP	(-) SP	(+) SP
$k_H\ (s^{-1})$	0.45	0.45	1.16§	1.16§
$k_B\ (s^{-1})$	430	430	430	430
$k_{R1}\ (s^{-1})$	0.090	13þ	0.090	13þ
$k_{R2}\ (s^{-1})$	0.081	13þ	0.081	13þ
Most of the values in the table come directly from the model specific fits presented in Figure 2.10.B. The exceptions are: (§) the stochastic model hydrolysis micro-constants have been adjusted by a factor of 2.59 to equate the mean complex lifetimes of the two systems. (þ) The micro-constant of ADP release in the presence of SP was taken from the value experimentally determined in Ref. 10.				

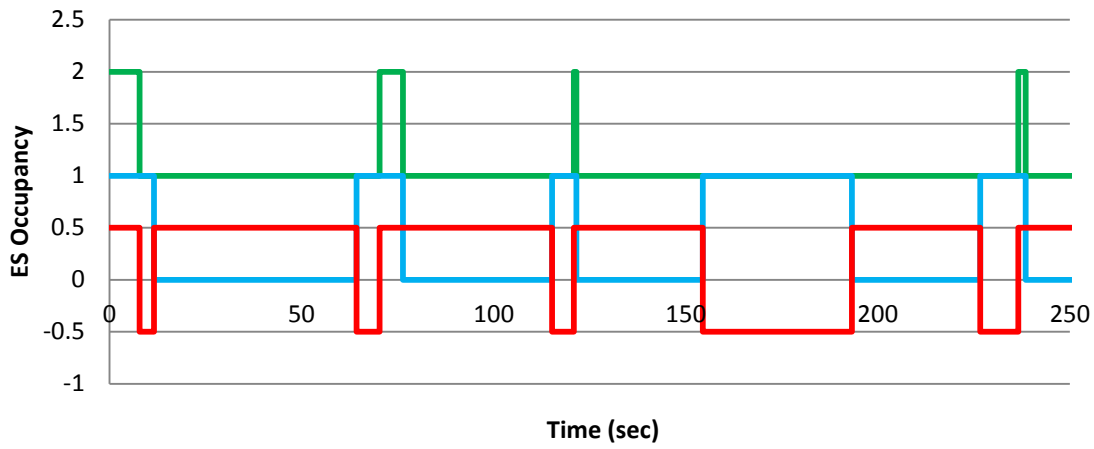
restated in Table 3.1. The results of the analysis, an average of 1000 single molecule simulations, is presented in Figure 3.9. The simulated set of data is in good agreement with the data, containing the same three phases (lag, burst and, delay) that were experimentally observed. This indicates that, when averaged over a large number of trials, the single molecule simulations are able to reproduce the results of the deterministic simulations.

The next set of single molecule simulations were performed on two different models, one corresponding to the concerted hydrolysis system, and the other to the stochastic hydrolysis system. The two sets of micro-constant values that were used are given in Table 3.1. The values for the micro-constants used were derived from the model specific fit results of Figure 2.11 and Table 2.3.B. The hydrolysis micro-constant in the stochastic hydrolysis system was adjusted by the factor 363/140, so that the lifetimes of a full ring in both systems were identical.

Under conditions of small values of the release micro-constant (absence of SP) the two systems behave very similarly. The release of nucleotide is rate-limiting, and so the asymmetric cycle is dominant. This is demonstrated by an out of phase exchange of GroES between the two rings, Figure 3.10. The resting state complex is dominant in both systems, as was predicted by the steady-state analysis and consistent with experiment. Finally, the distributions of GroEL/GroES complex lifetimes are near identical in both systems, dominated by the slow release of ADP instead of the hydrolysis mechanism itself.

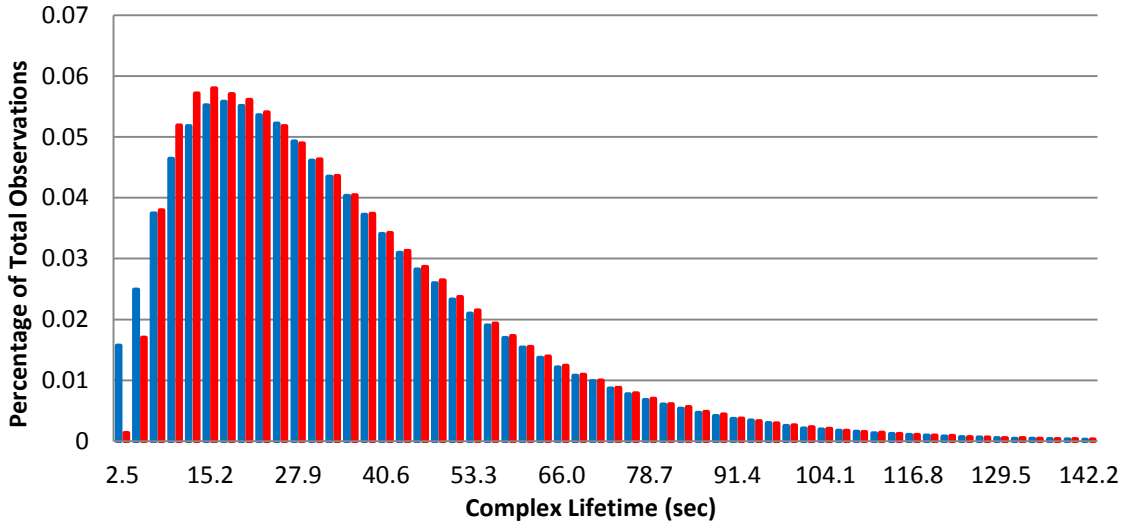
Figure 3.10 – Results of the Single Molecule Simulates at Low R Values (Absence of SP)

(A) Occupancy of Each GroEL Ring by GroES



(A) Single molecule simulated occupancy of **Ring A** and **Ring B** of GroEL by GroES. The trace presented was obtained from the concerted system, but is well representative of the stochastic system under conditions of low R. The graph takes the value of 1 when GroES is present and 0 when it is absent. The entire curve for **Ring A** has been offset by 0.5 to remove overlap and allow one to more easily compare the occupancy of the two rings. The **total number of GroES occupying the complex** is also given. Football complexes do have a propensity to form. Although not impossible, their lifetimes are extremely brief.

(B) Distribution of GroEL/GroES Complex Lifetimes



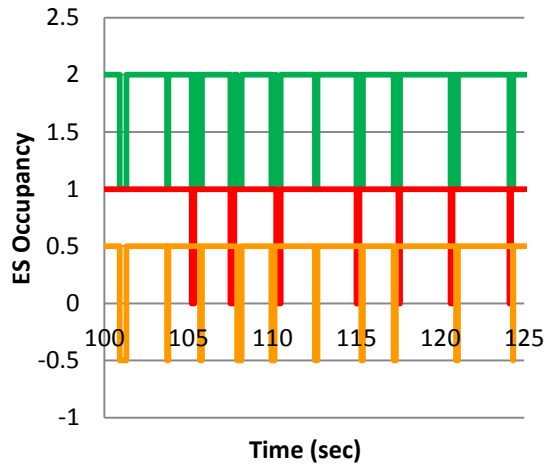
(B) Distribution of GroEL/GroES complex lifetimes obtained from single molecule simulations at low R values (absence of SP). The simulations were performed in order to compare the **concerted** and **stochastic** hydrolysis mechanisms. At low R, the distribution of complex lifetimes is nearly identical for both hydrolysis mechanisms.

Under conditions of high R, the two systems begin to exhibit distinctive characteristics. In both systems, the symmetric GroEL/GroES$_2$ species is dominant, Figure 3.11. However, the exchange behavior of the symmetric particles in the system is dependent on the system's hydrolysis mechanism. The simulation predict that the concerted mechanism of hydrolysis would lead to more "mis-cycles": events where the formation and dissociation of a GroEL$_7$/GroES complex

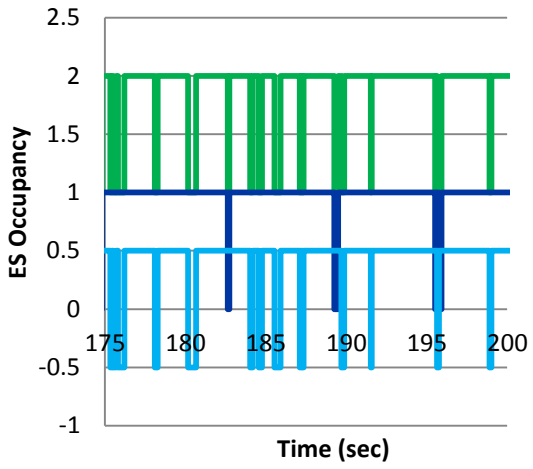
Figure 3.11 – Results of the Single Molecule Simulates at High R Values (Presence of SP)

Occupancy of Each GroEL Ring by GroES

(A) Stochastic Mechanism



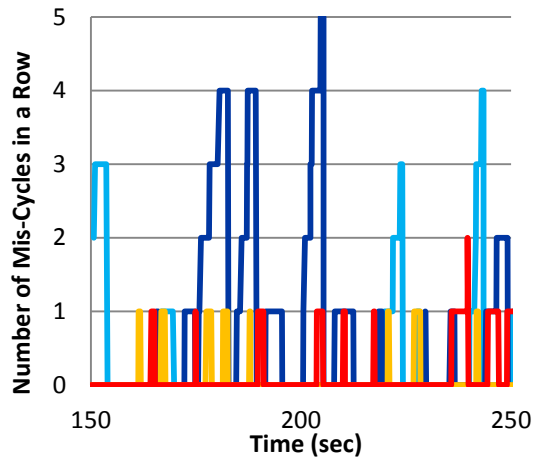
(B) Concerted Mechanism



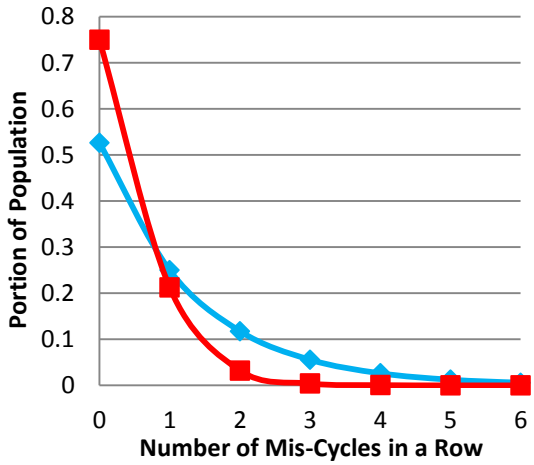
(A) Single molecule simulated occupancy of **Ring A** and **Ring B** of GroEL by GroES for the **stochastic hydrolysis** mechanism. Note the plot for **Ring B** has been offset 0.5 units for the sake of clarity (B) Single molecule simulated occupancy of **Ring A** and **Ring B** of GroEL by GroES for the **concerted hydrolysis** mechanism. Note the plot for **Ring B** has been offset 0.5 units for the sake of clarity. The graphs take the value of 1 when GroES is present and 0 when it is absent. The **total number of GroES occupying the complex** is also given.

Comparison of Ring Mis-Cycling by the Concerted and Stochastic Hydrolysis Mechanisms

(C) Number of Mis-Cycles Observed



(D) Propensity of Mis-Cycling



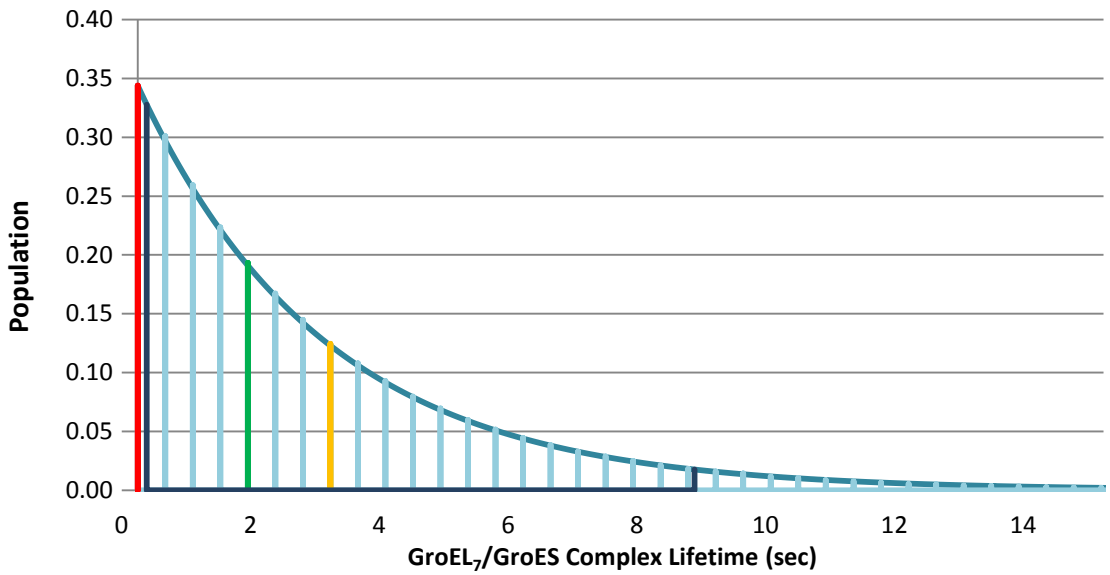
(C) The number of mis-cycling events observed in simulations of the two hydrolysis mechanisms in a randomly selected 100 second time period. The light and dark blue curves are **Ring A** and **Ring B** of the **concerted hydrolysis** system whereas the red and orange curves are **Ring A** and **Ring B** of the **stochastic hydrolysis** system.

(D) The overall probability of mis-cycling in the **concerted** vs **stochastic** hydrolysis systems simulated. Note that zero mis-cycles correspond to a regular turnover event. The probability of mis-cycling in the two systems was calculated by summing the number of mis-cycling events observed over 1000 x 500 second simulations.

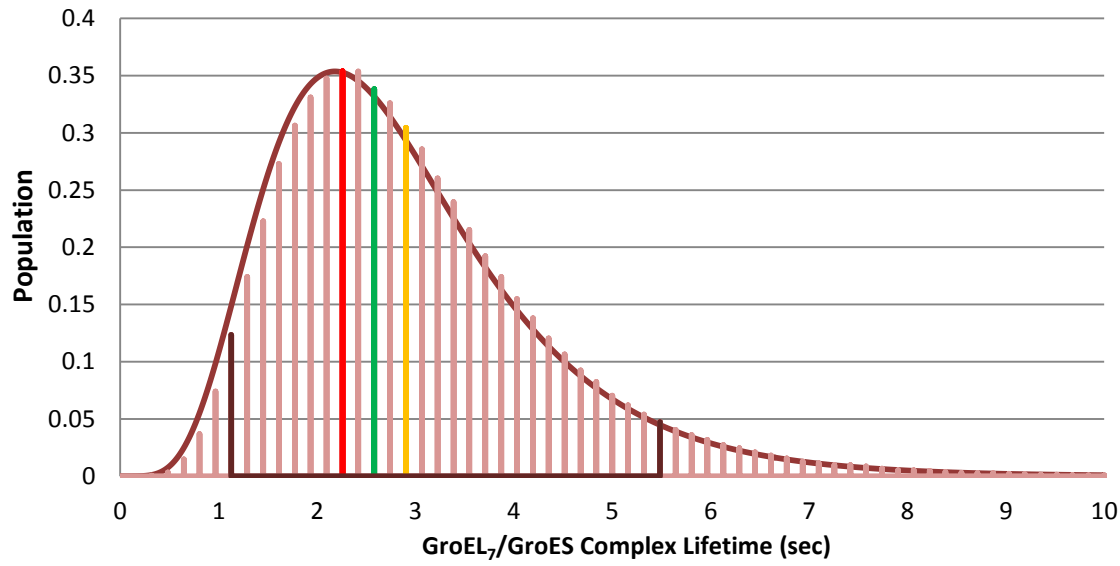
takes place on one ring, while the adjacent ring has not yet turned over. That is, a mis-cycle is when one ring turns over while the other is stuck as a $GroEL_7$/GroES complex. Figure 3.11 shows a plot of the number of mis-cycles for each ring in the two systems.

Figure 3.12 – Single Molecule Simulated Distribution of $GroEL_7/GroES$ Complex Lifetimes at High R (Presence of SP)

(A) Concerted ATP Hydrolysis Model



(B) Stochastic ATP Hydrolysis Model



Distributions of $GroEL_7/GroES$ complex lifetimes for the indicated hydrolysis model were calculated from observations of 1000 x 500 sec single molecule simulations. The micro-constant values used can be found in Table 3.1. For each system, the total range of observed lifetime values was evenly binned and the size of the population in each bin is plotted as the lightest color bar graph. The theoretical distribution of full ring lifetimes (Section 3.5.1) are the darker curves presented in each graph, and trace the experimental distributions well. The **mean**, **median** and, **mode** of each distribution is indicated. The darkest bars plotted indicate the 90% confidence interval of the distribution.

In the concerted system, it was observed that half of the $GroEL/GroES_2$ complexes formed will mis-cycle. This phenomenon can be directly attributed to the single-step concerted mechanism of ATP hydrolysis. The only symmetric species that can exist is the $(ATP)_7/(ATP)_7$ particle. The

exponential decay process governing the hydrolysis of ATP is “memory-less”. That is, when the $(ATP)_7/(ATP)_7$ species forms, both rings are equally likely to hydrolyze the bound nucleotide. In the concerted system, there is no way for the symmetric GroEL/$GroES_2$ particle to distinguish between the newly formed $GroEL_7$/GroES complex, and the one that existed in the previous round of hydrolysis. This “memory-less” property of the concerted system means it is more error prone than the stochastic system, leading to double the number of mis-cycling events.

Since hydrolysis is the rate-limiting step under conditions where the value of the release micro-constant is large (presence of SP), the behavior of the systems more closely reflects the behavior of these states. The second significant difference between the two hydrolysis mechanisms is the distribution of lifetimes of the $GroEL_7$/GroES complex (i.e. the SP encapsulation time) simulated, Figure 3.12. Despite the two systems having the same mean lifetime, the distributions are dramatically different. Symmetric complex lifetimes in the concerted system are distributed exponentially, whereas the distribution of lifetimes in the stochastic system is relatively normal with a slight positive skew. These distributions correspond nicely to the distributions predicted for the lifetime of the full ring previously established (curves in Figure 3.12). The mean, median, mode and 90% data range for each distribution are given in Figure 3.12.

3.6 CONCLUSION

3.6.1 Establishment of the Stochastic Hydrolysis Mechanism

The analysis presented in this chapter built off the results established in previous chapter to conclude that hydrolysis of ATP in the GroEL/GroES system is a stochastic process. The argument was based on the result, thoroughly vetted in Chapter 2, that the accumulation of inorganic phosphate in the system is well described as a single-step process. The FRET measurements used to monitor the concentration of the $GroEL_7$/GroES complex that were presented in this chapter established the existence of multiple symmetric GroEL/$GroES_2$ species

that preceded the release of GroES from the symmetric complex. The kinetic behavior of the D398A mutant was used to establish that the release of GroES from the symmetric complex is strongly coupled to the hydrolysis of ATP. This allowed us to conclusively identify the set of symmetric species observed as a series of hydrolysis states.

The question then turned to how a set of multiple hydrolysis states can release product in a single exponential fashion. A simple linear system was used to demonstrate that the special relationship between rate constants in a stochastic process lead to a single exponential cumulative function. Despite the fact that the exact analytical relationship does not hold in the case of cyclic systems, the substitution of stochastic hydrolysis for the concerted hydrolysis in our model specific fitting produced equivalent results. A stochastic hydrolysis mechanism is then established as the logical alternative to the concerted hydrolysis mechanism, accounting for all the observed behavior attributed to the concerted mechanism, as well as the observation of multiple states in the decay of the symmetric particle.

An additional pillar of support for the stochastic hydrolysis model stems from its ability to describe the behavior of the system under conditions where the concerted hydrolysis model failed: the presence of SP. In the presence of SP, the hydrolysis portion of the catalytic cycle becomes rate-determining. The stochastic model not only accounted for the steady-state behavior of the system under these conditions, but also a distinctive pre-steady state feature unique to the stochastic mechanism.

3.6.2 Theoretical Comparison of Stochastic and Concerted Hydrolysis

The last half of the chapter was dedicated to a comparison of the two mechanisms of hydrolysis. The stochastic mechanism of hydrolysis was shown to lead to a $GroEL_7$/GroES complex lifetime that was ~ 2.6 times longer, for a given hydrolysis micro-constant value. When this difference is accounted for, the two mechanisms produce the same steady state concentrations of the key

catalytic species over a wide range of ATP/ES binding and ADP release micro-constant values, indicating that they are equally well suited to respond to cellular demands for folding catalysis.

The largest difference between the two mechanisms was illustrated by the single molecule experiments performed at high ADP release micro-constant values, corresponding the presence of SP in the system. The concerted model was shown to be more error prone than the stochastic model, as demonstrated by the frequency and severity of mis-cycling events observed. This is due to the "memory-less" property of the mechanism that prevents it from being able to distinguish between the two rings upon symmetric complex formation. The distribution of encapsulation lifetimes also differed substantially between the two hydrolysis mechanisms. The concerted model of hydrolysis produced an exponential distribution of $GroEL_7$/GroES complex lifetimes, while the stochastic model produced a skewed normal distribution. Even without the biochemical evidence pointing towards the stochastic hydrolysis mechanism, these results suggest that the concerted mechanism of ATP hydrolysis may not be evolutionarily favorable.

The results suggest that concerted hydrolysis would be a less reliable "timer" mechanism for the encapsulation of SP. The 90% confidence interval of the exponential distribution of lifetimes observed is from 0.15 to 8.7 seconds, a greater than 50 fold difference. This is in contrast to the 5 fold difference observed for the 90% confidence interval of the stochastic hydrolysis mechanism. Although the mean of the two distributions is the same (by design because we adjusted the hydrolysis micro-constant) at 2.9 seconds, the most probable lifetime in the concerted model, i.e. the mode of the distribution, is 0.0 seconds, in contrast to 2.3 seconds observed for the stochastic model. Thus, an encapsulated SP would be potentially subject to a wider distribution of encapsulation times in the concerted hydrolysis model than if hydrolysis proceeded stochastically.

The results of the single molecule simulations are striking in part because they illustrate nature's uncanny ability to produce order from disorder. The stochastic hydrolysis model is, by definition,

a disordered hydrolysis process, by which subunits hydrolyzes ATP independently of one another. Nonetheless, as a timing mechanism, this model produces a distribution of complex lifetimes that is both more narrowly and more normally distributed than the concerted hydrolysis model. That is, the behavior of the GroEL/GroES complex is more predictable and reliable when ATP hydrolysis is modeled as a disordered, random process.

3.7 REFERENCES

1. Yifrach, Ofer and A. Horovitz. "Nested Cooperativity in the Atpase Activity of the Oligomeric Chaperonin Groel." *Biochemistry* 34, no. 16 (1995): 5303.

2. Cui, Q. and M. Karplus. "Allostery and Cooperativity Revisited." *Protein Sci* 17, no. 8 (2008): 1295-307.

3. Gresham, J.S. "Allostery and Groel: Exploring the Tenets of Nested Cooperativity." (PhD Dissertation) University of Maryland, 2004.

4. Gray, T. E. and A. R. Fersht. "Cooperativity in Atp Hydrolysis by Groel Is Increased by Groes." *FEBS Lett* 292, no. 1-2 (1991): 254-8.

5. Rye, H. S., S. G. Burston, W. A. Fenton, J. M. Beechem, Z. Xu, P. B. Sigler and A. L. Horwich. "Distinct Actions of Cis and Trans Atp within the Double Ring of the Chaperonin Groel." *Nature* 388, no. 6644 (1997): 792-8.

6. Ranson, N. A., G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich and H. R. Saibil. "Atp-Bound States of Groel Captured by Cryo-Electron Microscopy." *Cell* 107, no. 7 (2001): 869-79.

7. Ranson, N. A., N. J. Dunster, S. G. Burston and A. R. Clarke. "Chaperonins Can Catalyse the Reversal of Early Aggregation Steps When a Protein Misfolds." *J Mol Biol* 250, no. 5 (1995): 581-6.

8. Gorovits, B. M., J. Ybarra and P. M. Horowitz. "Atp Hydrolysis Is Critical for Induction of Conformational Changes in Groel That Expose Hydrophobic Surfaces." *J Biol Chem* 272, no. 11 (1997): 6842-5.

9. Yang, D., X. Ye and G. H. Lorimer. "Symmetric Groel:Groes2 Complexes Are the Protein-Folding Functional Form of the Chaperonin Nanomachine." *Proc Natl Acad Sci U S A* 110, no. 46 (2013): E4298-305.

10. Ye, X. and G. H. Lorimer. "Substrate Protein Switches Groe Chaperonins from Asymmetric to Symmetric Cycling by Catalyzing Nucleotide Exchange." *Proc Natl Acad Sci U S A* 110, no. 46 (2013): E4289-97.

Chapter 4

Breakage of GroEL/GroES$_2$ Complex Symmetry

4.1 INTRODUCTION

4.1.1 Background

Until recently, the role of the symmetric GroEL/$GroES_2$ complex in SP folding catalysis was not well understood. The symmetric complex had been observed in numerous biochemical experiments, as well as imaged using electron-microscopy, dating back as early as the mid-1990s (1-6). But, despite knowledge of its existence, researchers could not agree on the role, if any, played by the symmetric complex in catalysis. The two-stroke mechanism of GroEL/GroES function, which specifies the asymmetric $GroEL_{14}$/GroES complex as the functional unit, was backed by numerous structural, kinetic and, equilibrium studies and, as such, had widespread support throughout scientific community (7-11). Furthermore, the conditions that resulted in the presence of a population of symmetric complexes were not well understood, leading many to write off its existence as an experimental artifact.

Insurrection against the two-stroke model began with a series of Japanese publications (12-14). The researchers were able to demonstrate that the population of the symmetric GroEL/$GroES_2$ complex was dependent on the concentration of SP present in the reaction mixture. This led to the proposed two-cycle model (14, 15), the expanded version of which can be found in Figure 2.13. In the absence of SP, ADP release is the rate limiting step and the system resides in the asymmetric, two-stroke, cycle. In the presence of SP, the rate of ADP is increased by the binding of SP and hydrolysis becomes the rate limiting step, pushing the system into the symmetric cycle. The two-cycle model identifies the symmetric GroEL/$GroES_2$ complex as the key functional unit in SP folding catalysis.

The analysis in the previous chapter led to the surprising result that ATP hydrolysis in the GroEL/GroES catalytic cycle is stochastic and not concerted as is often assumed. It was also demonstrated that, due to the stochastic mechanism of hydrolysis, the population of symmetric

complex exists as an ensemble of ATP/ADP configurations. In theoretical analysis that followed, it was assumed, in order to facilitate comparison with the concerted hydrolysis model, that hydrolysis of all seven ATPs within a ring was required before GroES could dissociate. However, the establishment of a stochastic hydrolysis calls into question this assumption, leading to the question: which nucleotide configurations lead to the dissociation of GroES from the symmetric GroEL/GroES$_2$ complex. In this chapter, the relationship between a symmetric complex's distribution of hydrolyzed nucleotides and the dissociation of GroES has been termed the "breakage of symmetry". Numerous studies, as well as the single molecule simulations in the previous chapter, have demonstrate that in the absence of SP, all seven ATP molecules in the *cis* ring are routinely hydrolyzed before ADP is released from the *trans* ring, allowing for the exchange of GroES without the accumulation of the symmetric complex (7-10). However, in the presence of SP, as is the case *in vivo*, the breakage of symmetry becomes the dominant mechanism for GroES exchange and plays a fundamental role in governing the cycling of GroEL/GroES catalysis.

4.1.2 Overview

The focus of this chapter is on the experimental and theoretical work that has been done to solve the breakage of symmetry problem for the GroEL/GroES catalytic cycle. The breakage of symmetry problem concerns the determination of the link between nucleotide hydrolysis and GroES dissociation within the GroEL/GroES$_2$ complex. The ideal solution to this problem would be a quantitative relationship between the hydrolyzed nucleotide distribution of a given species (X ATP remaining on ring A and Y ATP remaining on ring B) and the propensity of GroES to dissociate from one of the two rings. The distribution of hydrolyzed nucleotide within a complex is referred to throughout the chapter as the specie's "nucleotide state" and is denoted by the bracket notation: $\langle X, Y \rangle$.

The first step towards solving the breakage of symmetry problem is to further develop the kinetics of stochastic hydrolysis, quantitatively establishing a link between the populations of the various species within the system. These results are used to develop a kinetic model for the breakage of symmetry. Evaluation of this kinetic model against experimental data allows us to postulate the form of the solution to the breakage of symmetry problem as a simple one-to-one correspondence between a species nucleotide state, $\langle X, Y \rangle$, and the probability of GroES dissociation, $BoS(X, Y)$. The results are applied to the analysis of a simplified experimental system, comprised of only a small subset of the possible nucleotide states.

The specified form of the solution obtained can be used to show the breakage of GroEL/GroES$_2$ complex symmetry is time-independent. This allows for the conversion of a system's hydrolysis state functions from the time domain to the domain of average number of hydrolysis events. The change in independent variables reduces the number of parameters needed to accurately describe the breakage of complex symmetry. The time-independent model for the breakage of symmetry is quantitatively developed, followed by a presentation of the two experimental data sets used to check the model. Least-squares fitting is then used to determine the solution to the breakage of symmetry problem, $BoS(X, Y)$. Potential error in the proposed solution is analyzed by comparing a series of closely related alternative solutions. The results of this chapter provide new insight into the mechanism driving the cycling in GroEL/GroES protein folding catalysis.

4.2 METHODS

4.2.1 D398A Single-Ring Turnover Experiments

An experimental techniques to reduce the number of nucleotide states that need to be considered is to use single-ring turnover experiments with the D398A mutant, performed by Xiang Ye. These experiments exploit the mutant's very slow ATPase. D398A is incubated with GroES and just enough ATP to populate a single ring, forming asymmetric complexes that contain 7 ATP

molecules in the *cis* ring, and none on the *trans* ring. The achievement of this homogeneous distribution of nucleotide states (all ⟨7,0⟩) is aided by the negative inter-ring cooperativity of ATP binding to GroEL. After the formation of the asymmetric complex, the mutant will begin to slowly hydrolyze the ATP to ADP. At various time points, the complexes are challenged with an excess of ATP and GroES, which populate the *trans* ring of the complex with GroES and seven ATP. GroES dissociation from the *cis* ring will only be observed once sufficient breakage of symmetry has been developed. For each time point, the distribution of nucleotide states in the *cis* ring can be back calculated using the measured rate of hydrolysis of the D398A mutant and assuming that the hydrolysis of ATP is stochastic.

4.2.2 Pre-Steady State Kinetics

One of the experimental methods used to address the breakage of symmetry problem is the pre-steady state kinetic experiments, performed by Xiang Ye, previously presented in Chapters 2 and 3. Specifically, the plots presented in Figure 3.4, which give the decay of the symmetric complex population as a function of the number of hydrolysis events for wild-type and D398A GroEL. The decay profile of the population of symmetric complexes will be used to fit the solution to the breakage of symmetry problem.

4.2.3 Subunit Mixing Experiments

In the presence of 5M urea, GroEL dissociates into monomers. These monomers can be reconstituted into active 14mers by dilution into a high salt buffer containing nucleotide. This technique allows for the mixture of subunits with differing sequences to be combined into a single 14mer complex. Subunit mixing was performed on two sets of two GroEL variants. The first, wild-type and K242C-TMR, were mixed to demonstrate the effectiveness of the procedure at producing a mixture of subunits. The results of this analysis can be found in Appendix A.4.3. The second, wild-type and D398A, prepared by Xiang Ye, were mixed to help address the

breakage of symmetry problem. The exact details of the subunit mixing procedure can be found in Appendix A.4.2.

After mixing, the wild-type/ D398A mixtures were labeled with donor fluorophore with the same procedure used for the pre-steady state FRET experiments presented in Chapter 3. The mixed subunit samples were challenged with an excess of acceptor labeled GroES and ATP, also as described in Chapter 3.

4.3 ANALYSIS

4.3.1 Stochastic Hydrolysis Kinetic Scheme

The realization that the mechanism of ATP hydrolysis in the GroEL/GroES system is stochastic has led us to question the relationship between the nucleotide state of the symmetric complex and the dissociation of GroES. Before we can address the link between the two processes, it will be helpful to develop the kinetic schemes, as well as the solutions to their state and product accumulation functions, that describe the stochastic hydrolysis of ATP in the complex. Two equivalent ways of formulating the kinetics of ATP hydrolysis are used throughout the chapter and developed below.

4.3.1.1 Hydrolysis State Formulation

The first formulation considered is a linear stochastic kinetic scheme that represents the GroEL 14mer as a single hydrolyzing unit. Hydrolysis events are treated as independent events that occur with equal probability (i.e. stochastic), which includes the implied independence of the two rings with respect to hydrolysis (not binding). GroEL can bind a total of 14 ATP molecules, indicating that 15 states, as well as the hydrolysis micro-constant H, are necessary to describe the kinetic scheme. The linear stochastic kinetic scheme that describes the ATPase of the system in this formulation is presented in Figure 4.1. The states of the scheme in Figure 4.1 correspond to the

Figure 4.1 – The Hydrolysis State Representation of the ATPase Kinetics of the GroEL/GroES System

$$\mathcal{Z}_0 \xrightarrow{14\,H} \mathcal{Z}_1 \xrightarrow{13\,H} \mathcal{Z}_2 \xrightarrow{12\,H} \cdots \mathcal{Z}_M \xrightarrow{(14-M)H} \mathcal{Z}_{M+1} \cdots \xrightarrow{H} \mathcal{Z}_{14}$$

The kinetic scheme above can be used to describe the hydrolysis kinetics of the GroEL/GroES system. The states of the system, $\mathcal{Z}_M$, correspond to the population of species that have undergone M hydrolysis events. Since the model is just a linear stochastic kinetic scheme, the analytic solution to the systems state functions and its product accumulation function can be analytically determined under certain conditions (see Section 4.3.1).

various "hydrolysis states" of the system: $\{\mathcal{Z}_M\}_{M=0}^{14}$, where $\mathcal{Z}_M$ is the total population of species that have undergone M hydrolysis events. The benefit of describing the ATPase activity of the GroEL/GroES system in terms of its hydrolysis states is due to the simplicity of the linear stochastic kinetic scheme. Under certain conditions, the product accumulation and state functions of the system can be easily determined.

Consider a linear system of $N+1$ states $\{\Psi_i\}_{i=0}^{N}$ (note the states are indexed starting with $i=0$), with initial conditions $\Psi_0 = 1$ and $\Psi_j = 0\ \forall\ j \in [1, N]$, that are connected by N stochastic processes, characterized by the rate constant k. It was shown in Section 3.3.1 that the product accumulation function of the system, $\Phi(t)$, is described by a single exponential accumulation function, Eq 4.1. The system of ODE equations corresponding to a stochastic system can be solved by the inverse Laplace method (Appendix A.2.4) to determine the general form of its state functions, Eq 4.2. Applying these results to the GroEL/GroES system, as represented in Figure 4.1, the hydrolysis states can be shown to be equivalent to the binomial distribution function with $N = 14$ and the probability term equal to $1 - e^{-H\,t}$, where H is the value of the hydrolysis micro-constant, Eq 4.3.

$$\Phi(t) = N - Ne^{-k\,t} \qquad \text{Eq 4.1}$$

$$\Psi_i = \frac{N!}{(N-i)!\ i!}\left(e^{-k\,t}\right)^{N-i}\left(1 - e^{-k\,t}\right)^{i} \qquad \text{Eq 4.2}$$

$$\mathcal{Z}_M = \frac{14!}{(14-M)!\ M!}\ (e^{-H\,t})^{14-M}(1-e^{-H\,t})^M \qquad \text{Eq 4.3}$$

4.3.1.2 Nucleotide State Formulation

The second way of formulating the stochastic hydrolysis kinetic scheme is to consider the behavior of each ring individually. This produces the two dimensional kinetic scheme of Figure 4.2. The individual rows correspond to linear stochastic kinetic schemes describing the hydrolysis in one ring, while the individual columns correspond to linear stochastic kinetic schemes describing hydrolysis in the adjacent ring. The states of scheme in Figure 4.2 are the so called "nucleotide states" of the system, $\{\langle X,Y\rangle\}_{X,Y=0}^{7}$, where $\langle X,Y\rangle$ is the total population of species containing X ATP on ring A and Y ATP on ring B.

The nucleotide state formulation of the ATPase activity of GroEL/GroES provides more detail than doing so in terms of its hydrolysis states. Close inspection of the two shows that while there are 15 hydrolysis states, there are 64 nucleotide states, 36 of which are unique for a symmetric species. The drawback is the loss of simplicity of the solution to the kinetic scheme, and the functional form of the states is no longer easily obtained. However, the two descriptions of the system are equivalent, and the relationship between the states of the two schemes is given in Eq 4.4.

Figure 4.2 – The Nucleotide State Representation of the ATPase Kinetics of the GroEL/GroES System

$$\begin{array}{ccccccccc}
\langle 7,7\rangle & \overset{7H}{\rightarrow} & \langle 7,6\rangle & \overset{6H}{\rightarrow} & \langle 7,5\rangle & \ldots & \langle 7,1\rangle & \overset{H}{\rightarrow} & \langle 7,0\rangle \\
\downarrow{\scriptstyle 7H} & & \downarrow{\scriptstyle 7H} & & \downarrow{\scriptstyle 7H} & & \downarrow{\scriptstyle 7H} & & \downarrow{\scriptstyle 7H} \\
\langle 6,7\rangle & \overset{7H}{\rightarrow} & \langle 6,6\rangle & \overset{6H}{\rightarrow} & \langle 6,5\rangle & \ldots & \langle 6,1\rangle & \overset{H}{\rightarrow} & \langle 6,0\rangle \\
\downarrow{\scriptstyle 6H} & & \downarrow{\scriptstyle 6H} & & \downarrow{\scriptstyle 6H} & & \downarrow{\scriptstyle 6H} & & \downarrow{\scriptstyle 6H} \\
\langle 5,7\rangle & \overset{7H}{\rightarrow} & \langle 5,6\rangle & \overset{6H}{\rightarrow} & \langle 5,5\rangle & \ldots & \langle 5,1\rangle & \overset{H}{\rightarrow} & \langle 5,0\rangle \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
\langle 1,7\rangle & \overset{7H}{\rightarrow} & \langle 1,6\rangle & \overset{6H}{\rightarrow} & \langle 1,5\rangle & \ldots & \langle 1,1\rangle & \overset{H}{\rightarrow} & \langle 1,0\rangle \\
\downarrow{\scriptstyle H} & & \downarrow{\scriptstyle H} & & \downarrow{\scriptstyle H} & & \downarrow{\scriptstyle H} & & \downarrow{\scriptstyle H} \\
\langle 0,7\rangle & \overset{7H}{\rightarrow} & \langle 0,6\rangle & \overset{6H}{\rightarrow} & \langle 0,5\rangle & \ldots & \langle 0,1\rangle & \overset{H}{\rightarrow} & \langle 0,0\rangle
\end{array}$$

The kinetic scheme above can be used to describe the hydrolysis kinetics of the GroEL/GroES system. The states of the system, $\langle X,Y\rangle$, correspond to the population of species that have X ATP on ring A and Y ATP on ring B.

$$\mathcal{Z}_M = \sum_{(7-X)+(7-Y)=M} \langle X,Y\rangle \qquad \text{Eq 4.4}$$

4.3.2 Time-Dependent Breakage of Symmetry Model

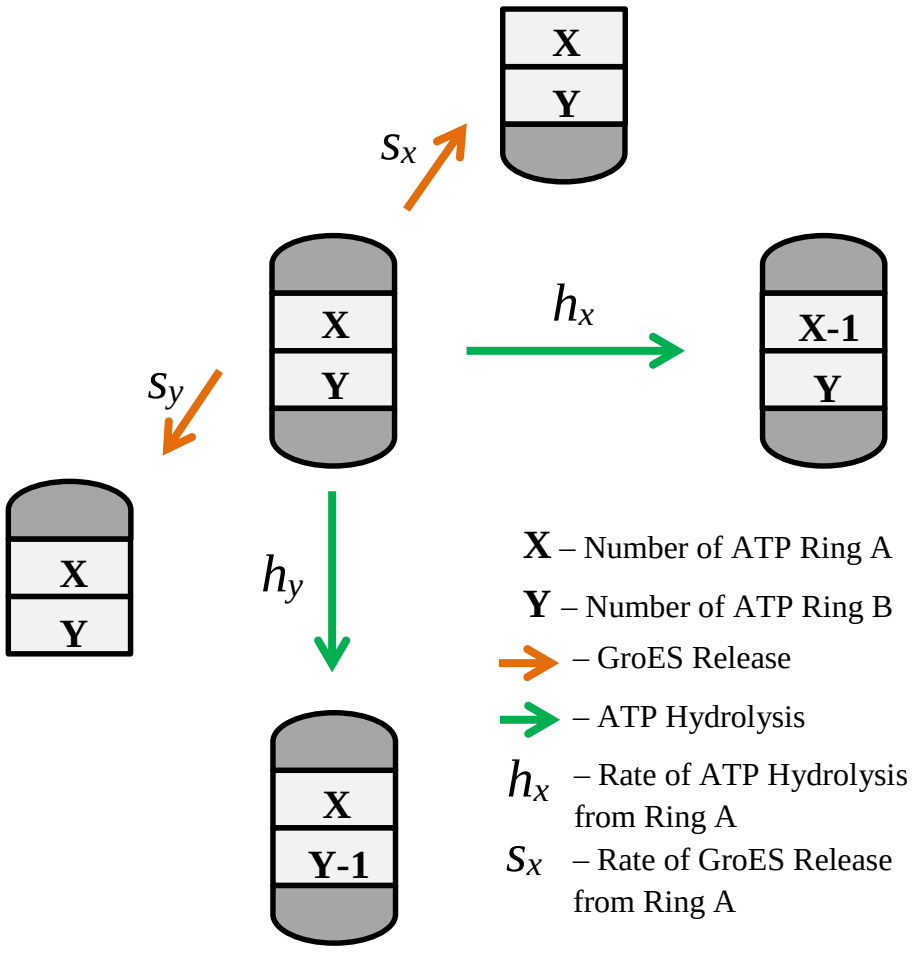
The time-dependent model of the breakage of symmetry was developed to address the problem in the time domain, which can be checked against kinetic experiments that monitor symmetric GroEL/GroES$_2$ complex formation and dissociation. The model was eventually rendered obsolete by kinetic analysis of the D398A mutant. Nonetheless, the "failure" of the model is actually very informative, demonstrating that the breakage of symmetry problem can be addressed in a time-independent manner. This lends valuable insight into the final form of the solution to the breakage of symmetry problem, $BoS(X,Y)$.

4.3.2.1 Theory

An illustration of the kinetic model of symmetry breakage is given in Figure 4.3. The model is simply an expansion of the nucleotide state formulation of the system's ATPase kinetics to include rate terms for the dissociation of GroES from each ring. Each nucleotide state of the symmetric complex is faced with four possible routes forward: hydrolyzing one molecule of ATP from ring A, hydrolyzing one molecule of ATP from ring B, releasing GroES from ring A or, releasing GroES from ring B. The probability of any one of the four processes occurring is dependent on the rate of that process relative to the three other paths forward: the kinetic competition between the four processes leads to the observed distribution of symmetric species.

The micro-constants of ATP hydrolysis are specified by the stochastic mechanism, such that a ring with X ATP has a micro-constant of $h_X = X\,H$. The value of H can be determined from fitting pre-steady state Pi accumulation data, as demonstrated in Chapter 2. The remaining challenge is then to determine the value of the various micro-constants that correspond to the rate of GroES dissociation from a ring, s_x. The relationship between micro-constants and the height of free energy barriers adds to the appeal of the model and can be used to make predictions

Figure 4.3 – Kinetic Model of Breakage of Complex Symmetry



The kinetic model of the breakage of GroEL/GroES$_2$ complex symmetry. The kinetic scheme of the model can be obtained by applying the modification illustrated above to each state of the nucleotide state formulation of the hydrolysis kinetic scheme. Each nucleotide state in the model would then feed into the two subsequent nucleotide states by stochastic hydrolysis, as well as an additional two states, corresponding to the loss of GroES from each ring. The model was rendered obsolete in light of experimentation on the D398A mutant.

regarding the relationship between the nucleotide state of the species and the GroES dissociation micro-constant of its rings.

Despite its theoretical appeal, the kinetic competition model was rendered obsolete by experimentation with D398A. Inherent in the kinetic competition model is the prediction that all nucleotide states have some non-zero probability of dissociating GroES. In fact, the probability of a particular nucleotide state, $\langle X, Y \rangle$, of the symmetric GroEL/GroES$_2$ complex dissociating GroES from ring A is given by Eq 4.5. Inspection of the equation shows that a change in the value of the

hydrolysis micro-constant, H, will cause a shift in the probability distribution of GroES dissociation for the entire set of nucleotide states of the symmetric complex, $\{\langle X,Y\rangle\}_{X,Y=0}^{7}$.

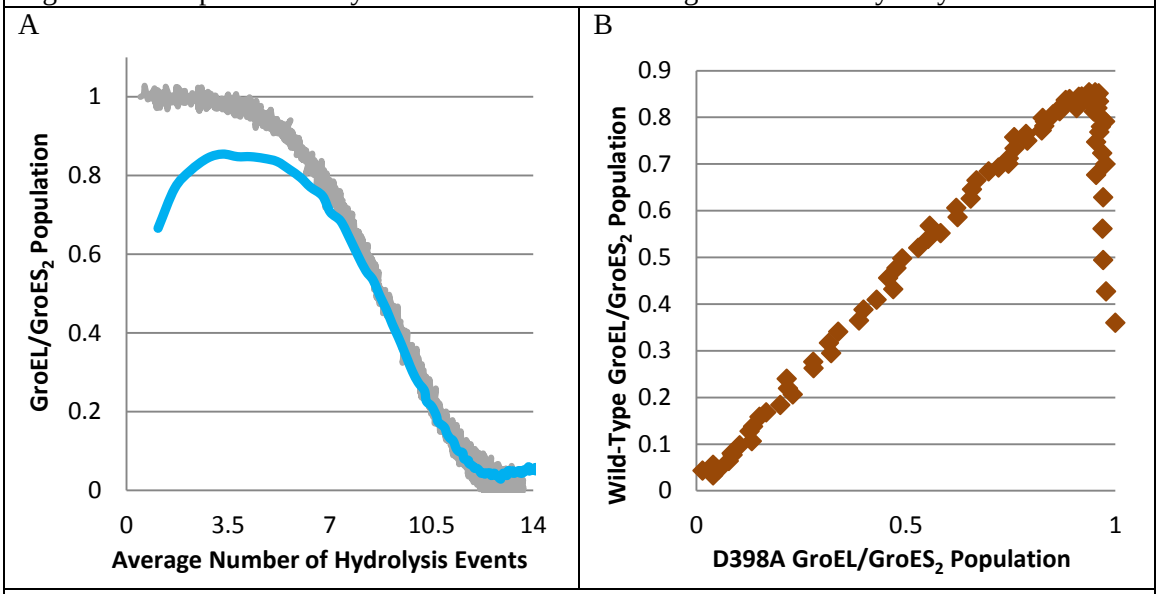
$$P_{ES}(X,Y) = \frac{s_X}{(X+Y)\,H + \,s_X + \,s_Y} \qquad \text{Eq 4.5}$$

4.3.2.2 Analysis

Figure 4.4.A contains the plots of the population of the symmetric particle against the average number of hydrolysis events that have occurred, i.e. $M = (7-X) + (7-Y)$, for both GroEL wild-type and the D398A mutant. As noted before, D398A is a GroEL variant that hydrolyzes ATP at ~0.1% of the wild-type rate. A 1000-fold decrease in the hydrolysis micro-constant is expected to be accompanied by a shift in the population of symmetric complexes, leading to a decrease in the population of symmetric complexes for a given nucleotide state. However, the results presented in Figure 4.4 demonstrate that no measureable change was observed in the underlying probability distribution of GroES dissociation, $P_{ES}(X,Y)$. Although the two curves do not identically trace one another, the differences can be shown to be attributed to the differing initial distribution of symmetric nucleotide states in the two systems.

Inspection of Figure 4.4.A shows that the plateau reached by the wild-type sample is not as high as the value reached by the D398A sample. This can be attributed to the competition between binding and hydrolysis, which introduces molecules into the population of symmetric particles after one or more hydrolysis events (i.e. GroEL/GroES$_2$ complexes form at nucleotide states other than ⟨7,7⟩). Nonetheless, in both systems the decay of the population of symmetric particles begins at the same number of hydrolysis events and continues along the same trajectory through the various hydrolysis states. This can be illustrated by a plot of the two experimentally determined curves against one another, Figure 4.4.B. The linear relationship between the two curves indicates that as a symmetric particle proceeds through each hydrolysis state, their journey is terminated (i.e. GroES dissociates and the complex is no longer a member of the population of

Figure 4.4 – Population of Symmetric Particles vs Average Number of Hydrolysis Events



(A) Decay of the symmetric GroEL/GroES$_2$ complex plotted against average number of hydrolysis events, obtained from pre-steady state kinetic measurements of GroEL$_7$/GroES association and Pi release, performed by Xiang Ye. Two GroEL variants were tested: **wild-type** and the slow ATP hydrolyzing **D398A mutant**. (B) The two plots in (A) do not overlay perfectly due to the competition between ATP binding and hydrolysis in the wild-type sample which is negligible in the limit of slow hydrolysis. Nonetheless, it can be demonstrated that **the two populations take the same path through the hydrolysis states** by simply plotting the **wild-type** and **D398A** curves in (A) against one another.

symmetric particles) at the same nucleotide states. In other words, at each hydrolysis state, the proportion of the population of symmetric particles lost (i.e. that dissociates GroES) is the same in both systems. This indicates that there was no change in the underlying probability distribution of GroES dissociation, $P_{ES}(X,Y)$, for any of the nucleotide states, $\langle X,Y \rangle$.

4.3.3 Form of Solution to the Breakage of Symmetry Problem

The results of the above analysis are quite informative with regard to the form of the solution to the breakage of symmetry problem, demonstrating that the values of the micro-constants governing the rates of GroES dissociation are quite special. The experiment indicates that the probability of GroES dissociation is insensitive to the value of the hydrolysis micro-constant. This implies that the derivative of P_{ES} with respect to H is zero for all values of X, Y and H. This condition is only met in the limit of extremely large or extremely small values of s_X. It follows that the dissociation of GroES from a nucleotide state $\langle X,Y \rangle$ occurs with either zero probability or probability equal to unity.

$$\frac{dP_{ES}}{dH}(X,Y) = 0 \quad \forall\, X,Y \in [0,7]$$

$$\frac{dP_{ES}}{dH}(X,Y) = \frac{-(X+Y)\, s_X}{\left((X+Y)\, H + s_X + s_Y\right)^2} = 0$$

$$Then\ \forall\ \langle X,Y\rangle \in \{\langle X,Y\rangle\}_{X,Y=0}^{7}, \qquad either \qquad s_X = 0\ or\ s_X \gg (X+Y)H + s_Y$$

$$which\ \ implies\ that\ \ \forall\ \langle X,Y\rangle \in \{\langle X,Y\rangle\}_{X,Y=0}^{7}, \quad P_{ES}(X,Y) = 0\ or\ 1$$

Since it has been established that $P_{ES}(X,Y) = 0\ or\ 1$ for all nucleotide states $\langle X,Y\rangle$, it seems appropriate to state a few other assumptions regarding the nature of the relationship between $\langle X,Y\rangle$ and $P_{ES}(X,Y)$ that are applied throughout the chapter. First, implicit in the model presented in Figure 4.3 is the assumption that the probability of GroES dissociation is independent of the topological variation within a given nucleotide state. That is, all the various unique arrangements of ATP/ADP that make up a given nucleotide state $\langle X,Y\rangle$ have the same probability of GroES dissociation, $P_{ES}(X,Y)$. There is no reason *ab initio* to believe that the breakage of symmetry is topologically invariant; rather, the assumption is implemented in accordance with Occam 's razor.

Three additional assumptions regarding the probability of GroES dissociation from a symmetric complex, $P_{ES}(X,Y)$, are routinely utilized throughout the chapter to reduce the complexity of the breakage of symmetry problem.

$$P_{ES}(X,Y) = P_{ES}(Y,X)$$

$$If\ X < Y\ and\ P_{ES}(X,Y) = 1, then\ P_{ES}(X,y) = 1\ \forall\ y > Y$$

$$If\ X < Y\ and\ P_{ES}(X,Y) = 1, then\ P_{ES}(x,Y) = 1\ \forall\ x < X$$

The first assumption is simply a reflection of the symmetry of a GroEL/GroES$_2$ complex. The other two assumptions seem to follow logically from the known stability of the $\langle 7,7\rangle$ complex

and the (as-of-yet unknown) structural mechanism that leads to the release of GroES from the symmetric complex (it seems unlikely that $\langle X+1, Y\rangle$ and $\langle X-1, Y\rangle$ would be stable symmetric complexes but not $\langle X, Y\rangle$). They also reflect the assumption that GroES dissociation occurs from the ring containing the fewer number of remaining ATP molecules.

The breakage of symmetry problem seeks to address which nucleotide states, $\langle X, Y\rangle$, lead to the dissociation of GroES from the symmetric complex. A comparison of the decay of the GroEL/GroES$_2$ complex of the wild-type and D398A GroEL variant led to the conclusion that the value of $P_{ES}(X, Y)$ is either 0 or 1, for all $X, Y \in [0,7]$. Since there are a finite number of possible nucleotide states, $\langle X, Y\rangle$, the solution to the breakage of symmetry problem, $BoS(X, Y)$, can be neatly represented as an 8 x 8 matrix, each entry being either zero or one. It should be noted that the matrices presented in this chapter are indexed by hydrolysis events, $[(7-X), (7-Y)]$, instead of number of ATPs in the nucleotide state $\langle X, Y\rangle$. That is, the upper left-most entry corresponds to $P_{ES}(7,7)$ and the lower right-most entry corresponds to $P_{ES}(0,0)$.

4.3.4 D398A Single Ring Turnover Experiments

4.3.4.1 Theory

The results of the previous section, specifying the form of $BoS(X, Y)$, can be utilized to analyze kinetic experiments that monitor the turnover of a single ring of the D398A variant of GroEL. The benefit of the D398A single-ring turnover experiments, performed by Xiang Ye, is that they address the breakage of symmetry problem within a reduced subset of the possible nucleotide states. Specifically, the subset of states of the form $\langle X, 7\rangle$ where $X \in [0, 7]$. The reduction of states greatly simplifies the problem and leads to a one-to-one correspondence between the system's hydrolysis states and its nucleotide states. Since the reduced subset of possible nucleotide states are connected by a linear stochastic kinetic scheme, it follows from application of Eq 4.2 that the state functions of the system are given by Eq 4.6.

$$\mathcal{X}_M(t) = \langle 7-M, 7\rangle = \frac{7!}{(7-M)!\ M!}\ (e^{-H\,t})^{7-M}(1-e^{-H\,t})^{M} \qquad \text{Eq 4.6}$$

Utilizing the assumptions stated in Section 4.3.3 above, it can be shown that the identification of a single cutoff value, w_7, is sufficient to describe the breakage of symmetry problem for the subset of states $\{\langle X, 7\rangle\}_{X=0}^{7}$. That is, if w_7 is the largest integer such that $P_{ES}(w_7, 7) = 1$, then $P_{ES}(X, 7) = 1\ \forall\ X \in [0, w_7]$ and $P_{ES}(X, 7) = 0\ \forall\ X \in [w_7 + 1, 7]$.

Assuming that asymmetric D398A/GroES system is challenged at time t with excess ATP and GroES, the function quantifying the population of particles that will dissociate is given by Eq 4.7. The value of hydrolysis micro-constant, H, for D398A can be determined by a fitting to its pre-steady state Pi accumulation data (Appendix A.4.1). This means that for all t and x the value of $F(t, x)$ can be evaluated.

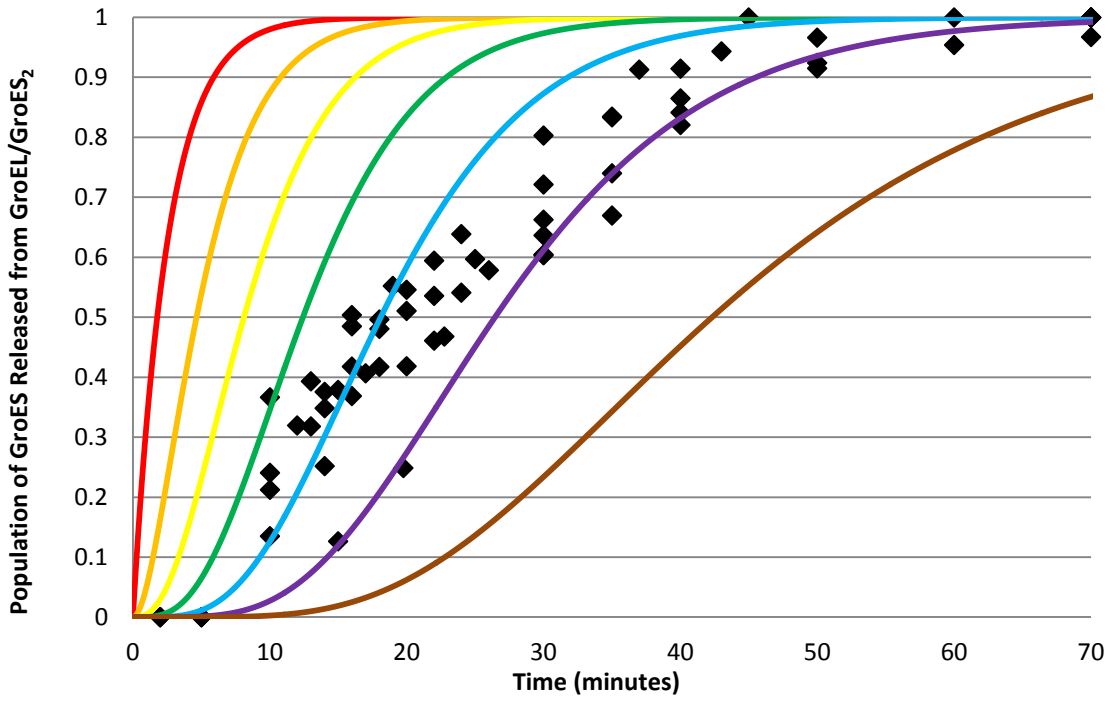
$$F(t, w_7) = \sum_{k=0}^{7} P_{ES}(7-k, 7)\ \mathcal{X}_k(t) = \sum_{k=7-w_7}^{7} \mathcal{X}_k(t) \qquad \text{Eq 4.7}$$

4.3.4.2 Results and Discussion

The results of the D398A single ring turnover experiments, obtained by Xiang Ye over a several trials, are plotted in Figure 4.5. The results are presented along with the family of curves, $\{F(t, j)\}_{j=0}^{7}$, which represents all possible cutoff values, w_7. The experimental data appears to trace the theoretical curve $F(t, 2)$ at early time points and tends towards $F(t, 1)$ at later time points. The drift of the data at later time points is assumed to be due to experimental error, supported by the spread of the data. The large experimental error is likely due to fault in the assumption that a homogenous asymmetric $\langle 0,7\rangle$ system was created at $t = 0$. Inspection of the curves in Figure 4.5 allow us to set the cutoff value $w_7 = 2$.

In order to solve the breakage of symmetry problem, we must specify the function $P_{ES}(X, Y)$ for all $X, Y\ \in [0,7]$. At first, it may seem like there are a wide number of possibilities for $P_{ES}(X, Y)$.

Figure 4.5 – D398A Single-Ring Turnover Experiments



The **experimental data** from the D398A single ring turnover experiments, performed by Xiang Ye, is presented along with the set of standard curves corresponding to **one**, **two**, **three**, **four**, **five**, **six** and, **seven** hydrolysis events required to break the symmetry of the $\langle X, 7\rangle$ complex.

However, the assumptions laid out in Section 4.3.2 can be used to show that a set of 8 parameters, $\{w_Y\}_{Y=0}^{7}$, where w_Y is the largest valued integer such that $P_{ES}(w_Y, Y) = 1$, is sufficient to uniquely determine the form of $P_{ES}(X, Y)$ for all $X, Y \in [0,7]$. In fact, setting $w_7 = 2$ drastically narrows the possible values of all other parameters in the set such that $w_Y \in \{0,1,2\}$ for $Y < 7$. Thus, analysis of an isolated subset of nucleotide states has allowed us to determine 31 out of 64 entries in the solution to the breakage of symmetry problem, $BoS(X, Y)$, Eq 4.8.

$$BoS(X,Y) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & \square & \square & \square \\ 0 & 0 & 0 & 0 & 0 & \square & \square & \square \\ 0 & 0 & 0 & 0 & 0 & \square & \square & \square \\ 0 & 0 & 0 & 0 & 0 & \square & \square & \square \\ 1 & \square & \square & \square & \square & \square & \square & \square \\ 1 & \square & \square & \square & \square & \square & \square & \square \\ 1 & \square & \square & \square & \square & \square & \square & \square \end{bmatrix} \qquad \text{Eq 4.8}$$

4.3.5 Time-Independent Breakage of Symmetry Model

Before the solution to the breakage of symmetry problem can be determined, we must first develop the quantitative relationship between the probability distribution of GroES dissociation, $P_{ES}(X,Y)$, and an experimentally obtainable data set. The analysis in Section 4.3.3 demonstrates the lack dependence of the solution to the breakage of symmetry problem on the value of the hydrolysis micro-constant, implying the solution is time-independent on the timescale considered. This finding introduces the opportunity of exploring the problem outside of the time domain. The following sections will demonstrate how the distribution of hydrolysis states can be reformulated as functions of another experimentally measurable independent variable: the average number of hydrolysis events. This conversion allows us to formulate a breakage of symmetry model that follows logically from the proposed stochastic hydrolysis mechanism and depends solely on the function $P_{ES}(X,Y)$. This model can be fit to experimentally obtained data sets, utilizing the above stated assumptions to narrow the possible outcomes, leading to a solution to the breakage of symmetry problem.

4.3.5.1 Development of Theory

4.3.5.1.1 Converting $\mathcal{Z}_M(t) \rightarrow \mathcal{Z}_M(\langle P \rangle)$

Looking again at the kinetic scheme in Figure 4.1 and utilizing the logic laid out in Section 4.3.2, the product accumulation function can be shown to be Eq 4.9, The equation can be rearranged and solved for $e^{-H\,t}$, Eq 4.10. Plugging this result into Eq 4.3, we obtain the hydrolysis state function within the domain of observed hydrolysis events, $\langle P \rangle$. The resulting formulation of the state function, Eq 4.11, is independent of time and the value of the micro-constant H. The newly derived independent variable, $\langle P \rangle$, is simply the experimentally determinable accumulation of inorganic phosphate.

$$P(t) = 14 - 14e^{-H\,t} = \langle P \rangle \qquad \text{Eq 4.9}$$

$$e^{-H\,t} = 1 - \frac{\langle P \rangle}{14} \qquad \text{Eq 4.10}$$

$$\mathcal{Z}_M(\langle P \rangle) = \frac{14!}{(14-M)!\ M!}\left(1 - \frac{\langle P \rangle}{14}\right)^{14-M}\left(\frac{\langle P \rangle}{14}\right)^{M} \qquad \text{Eq 4.11}$$

The size of the population of symmetric particles lost at a particular value of $\langle P \rangle$ is not thought to be due to entire population of species in $\mathcal{Z}_M(\langle P \rangle)$, but rather only the fraction of $\mathcal{Z}_M(\langle P \rangle)$ corresponding to the nucleotide state $\langle X, Y \rangle$ (where $X + Y = M$) such that $P_{ES}(X, Y) = 1$. Before the family of functions in Eq 4.11 can be modified to describe the decay of the symmetric particle on the domain of observed hydrolysis events, we must first develop the tools to quantify the proportion of a given hydrolysis state that exists as a symmetric particle. The procedure for accomplishing this will be presented in two steps. The first step is to determine the distribution of nucleotide states within a given hydrolysis state, which can be accomplished by the application of the stochastic hydrolysis mechanism. The next step is the incorporation of the probability distribution of GroES dissociation, $P_{ES}(X, Y)$, which allows for the quantification of the population of symmetric species as a function of observed hydrolysis events.

4.3.5.1.2 Decomposing a Hydrolysis State into Nucleotide States

As explained above, the first step towards quantifying the population of symmetric particles as a function of hydrolysis events is to determine the distribution of nucleotide states within a hydrolysis state. That is, for every $x, y \in [0,7]$, we seek the occupation coefficient, $\mathcal{U}_{x,y}$:

$$\mathcal{U}_{x,y} = \frac{\langle x, y \rangle}{\sum_{X+Y=x+y} \langle X, Y \rangle} = \frac{\langle x, y \rangle}{\mathcal{Z}_{(7-x)+(7-y)}}$$

Because then:

$$\mathcal{U}_{x,y} * \mathcal{Z}_{(7-x)+(7-y)} = \langle x, y \rangle$$

That is, given that a complex has undergone $(7 - x) + (7 - y)$ hydrolysis events, the value of $\mathcal{U}_{x,y}$ corresponds to the probability the complex is in nucleotide state $\langle x, y \rangle$. These occupation coefficients can be uniquely determined, for a given initial distribution of hydrolysis states, by application of the stochastic hydrolysis model.

The occupancy coefficients of each nucleotide state can be determined by $H_{Sto}(X, Y)$, Eq 4.12. $H_{Sto}(X, Y)$ is a recursively defined "function" that gives the probability of occupation of a nucleotide state $\langle X, Y \rangle$ that arises from the stochastic mechanism of hydrolysis, accounting for its dependence on the probability of occupancy of the preceding states $\langle X + 1, Y \rangle$ and $\langle X, Y + 1 \rangle$.

$$H_{Sto}(X, Y) = H_{Sto}(X + 1, Y)\frac{X + 1}{X + 1 + Y} + H_{Sto}(X, Y + 1)\frac{Y + 1}{Y + 1 + X} + H_O(X, Y) \qquad \text{Eq 4.12}$$

Because the "function" is recursively defined, it requires the input of seed values, $H_O(X, Y)$, corresponding to the initial distribution of nucleotide states in a system. When an initial distribution of nucleotide states is input into the "function", it generates the distribution of nucleotide states that arises from the stochastic hydrolysis mechanism connecting states. That is if $H_O(X, Y)$ is defined for every $X, Y \in [0,7]$, then $H_{Sto}(X, Y) = \mathcal{U}_{X,Y}$.

An example of a nucleotide state distribution, $H_{Sto}(X, Y)$, obtained by apply a seed values of $H_O(7,7) = 1$ and $H_O(X, Y) = 0$ for all other nucleotide states, is given in Figure 4.6.A. These conditions correspond to a system where the entire population of species initially occupies the nucleotide state $\langle 7,7 \rangle$. Each minor diagonal corresponds to a particular hydrolysis state, and its entries sum to unity. Also, note how the most probable nucleotide state in each hydrolysis state is the one with the highest nucleotide inter-ring nucleotide symmetry.

Figure 4.6 – Hypothetical Distributions of Nucleotide States Predicted by Stochastic Hydrolysis of ATP

(A) $H_{Sto}(X,Y)$

1	0.5	0.23	0.1	0.03	0.01	0	0
0.5	0.54	0.4	0.24	0.12	0.05	0.01	0
0.23	0.4	0.44	0.37	0.24	0.13	0.05	0.01
0.1	0.24	0.37	0.41	0.36	0.24	0.12	0.03
0.03	0.12	0.24	0.36	0.41	0.37	0.24	0.1
0.01	0.05	0.13	0.24	0.37	0.44	0.4	0.23
0	0.01	0.05	0.12	0.24	0.4	0.54	0.5
0	0	0.01	0.03	0.1	0.23	0.5	1

(B) $Q(X,Y)$

1	0.5	0.23	0.1	0	0	0	0
0.5	0.54	0.4	0.24	0.1	0	0	0
0.23	0.4	0.44	0.37	0.23	0.09	0	0
0.1	0.24	0.37	0.41	0.35	0.21	0	0
0	0.1	0.23	0.35	0.4	0.34	0.14	0
0	0	0.09	0.21	0.34	0.41	0.3	0
0	0	0	0	0.14	0.3	0.41	0
0	0	0	0	0	0	0	0

(C) $P_{ES}(X,Y)$

0	0	0	0	1	1	1	1
0	0	0	0	0	1	1	1
0	0	0	0	0	0	1	1
0	0	0	0	0	0	1	1
1	0	0	0	0	0	0	1
1	1	0	0	0	0	0	1
1	1	1	1	0	0	0	1
1	1	1	1	1	1	1	1

(A) Example of the distribution of nucleotide states that would be generated due to stochastic hydrolysis if ⟨7,7⟩ was the initial state of the system. (B) The distribution of symmetric nucleotide states in the same system can be obtained by inputting a $P_{ES}(X,Y)$. The example distribution shown uses the $P_{ES}(X,Y)$ shown in (C). Each nucleotide state along a minor diagonal corresponds to one hydrolysis state. **Boxed states** in (A) and (B) show how the distribution begins to change as symmetric particles are lost due to GroES dissociation. Note: the matrices are arranged so that ⟨7,7⟩ and ⟨0,0⟩ are the top left and bottom right entries, respectively.

4.3.5.1.3 Decomposing a Hydrolysis State into Symmetric Nucleotide States

Up until this point, we have only considered the relationship between nucleotide states that arise from the stochastic hydrolysis mechanism. We now wish to discriminate between nucleotide states of the symmetric and asymmetric complex. This too requires a recursively defined "function". Consider the following scenario: after $(7-X)+(7-Y)$ hydrolysis events, a symmetric particle reaches the nucleotide state $\langle X,Y\rangle$, which dissociates GroES (i.e. $P_{ES}(X,Y)=1$). Then the state $\langle X,Y\rangle$ no longer contributes to the population of symmetric species occupying nucleotide states $\langle X-1,Y\rangle$ or $\langle X,Y-1\rangle$ upon hydrolysis of ATP.

In order to account for the problem presented above, and accurately determine the size of the population of symmetric particles, the "functions" $H_{Sto}(X,Y)$ and

$P_{ES}(X,Y)$ are combined into a single recursively defined "function", $Q(X,Y)$, Eq 4.13. This equation, when seeded with a distribution of symmetric nucleotide states, $Q_O(X,Y)$, generates the subsequent occupancy of a nucleotide state $\langle X,Y\rangle$ by symmetric species only.

Figure 4.6.B contains an example distribution of symmetric nucleotide states, $Q(X,Y)$, obtained by seeding with the value $Q_O(7,7) = 1$ and $Q_O(X,Y) = 0$ for all other nucleotide states. The values of $Q(X,Y)$ were generated by considering a random $P_{ES}(X,Y)$ distribution, given in Figure 4.6.C. The removal of a nucleotide state from the population of symmetric species, which occurs when $P_{ES}(X,Y) = 1$ not only eliminates that state from the distribution given by Q, but also reduces the population of species downstream from the state. This can be seen by comparison of the blue boxed states in Figure 4.6A and 4.6.B.

$$Q(X,Y,Q_O,P_{ES}) = \left(Q(X+1,Y)\frac{X+1}{X+1+Y} + Q(X,Y+1)\frac{Y+1}{Y+1+X} + Q_O(X,Y)\right)(1 - P_{ES}(X,Y)\,) \qquad \text{Eq 4.13}$$

4.3.5.1.4 Conclusion

The results of the previous three subsections can be combined to obtain the desired quantitative relationship between experiment and the solution to the breakage of symmetry problem. The "function" $Q(X,Y,Q_O,P_{ES})$ can be used, along with Eq 4.11, to simulate the decay of the population of symmetric species as a function of the average number of hydrolysis events, $G(\langle P\rangle, Q_O, P_{ES})$ in Eq 4.14.

$$G(\langle P\rangle, Q_O, P_{ES}) = \sum_{i=0}^{7}\sum_{j=0}^{7} Q(i,j,Q_O,P_{ES})\, \mathcal{Z}_{(7-i)+(7-j)}(\langle P\rangle) \qquad \text{Eq 4.14}$$

$G(\langle P\rangle, Q_O, P_{ES})$ is independent of time, depending only on the average number of observed hydrolysis events, $\langle P\rangle$, the initial distribution of symmetric nucleotide states, $Q_O(X,Y)$, and the probability of GroES dissociation from a symmetric complex, $P_{ES}(X,Y)$. $\langle P\rangle$ is the independent parameter in Eq 4.14 and is experimentally determined as the measured accumulation of

inorganic phosphate. The initial distribution of symmetric nucleotide states, $Q_O(X,Y)$, must be known, however certain experimental conditions (discussed below) allow for $Q_O(X,Y)$ to be approximated with a high degree of accuracy. The final undetermined parameter in Eq 4.14, is $P_{ES}(X,Y)$. The various possible forms of $P_{ES}(X,Y)$, limited by the assumptions stated in Section 4.3.3, can be input into Eq 4.14 and fit to the experimentally determined data sets, providing a solution to the breakage of symmetry problem.

4.3.5.2 Experiments

The theory developed in the preceding section provides a quantitative link between the population of symmetric GroEL/GroES$_2$ complexes and the average number of hydrolysis events observed. Two different ways of experimentally generating data sets that correspond to the theory were performed by my colleague Xiang Ye. These data sets were fit to the theory outlined above in order to determine the solution to the breakage of symmetry problem.

4.3.5.2.1 FRET vs Pi Release

It was previously shown, in Chapter 2, that the average number of hydrolysis events can be followed experimentally by measuring the amount of Pi released. In Chapter 3, it was demonstrated that the population of GroEL$_7$/GroES complexes can be monitored experimentally using a FRET based system. These two experimental results can be plotted against one another to generate the type of curves presented in Figure 4.4.A, which correspond to the population of symmetric particles against the average number of hydrolysis events observed.

As noted, the above theory requires knowledge of the initial distribution of symmetric nucleotide states. In the case of the D398A mutant, the slow rate of ATP hydrolysis allows us to assume complete occupation of the symmetric nucleotide state ⟨7,7⟩ without introducing any significant error. This is not the case for wild-type GroEL, and knowledge of the initial distribution of nucleotide states would depend on the accuracy of fitted micro-constant values for ATP binding,

GroES binding and, ATP hydrolysis. The fitting performed in Section 4.3.4.3 to determine the solution to the breakage of symmetry problem excluded the FRET versus Pi data obtained from the wild-type system to avoid any unnecessary introduction of uncertainty in the input of $Q_O(X, Y)$.

4.3.5.2.2 Subunit Mixing

Numerous studies over the years have demonstrated the ability of urea to dissociate the $GroEL_{14}$ oligomer completely into its monomeric subunits and then reconstituted back into functional 14mers, Appendix A.4.2 (18-21). Appendix A.4.3 provides experimental validation that the process constitutes a mixing of subunits. Previously studies indicate that the total dissociation of the GroEL 14mers into individual subunits is complete at urea concentrations around 5M. In the analysis presented, it is assumed that if two different GroEL variants are mixed (say, wild-type and mutant GroEL), and the mixed sample is brought to 5M urea, inducing complete dissociation into monomers, reconstitution of the sample into functional 14mers will lead to a stochastic distribution of distinct subunits within the tetradecamer. Experimental evidence of this stochastic distribution, obtained by Dr. Justin Benesch's lab at the University of Oxford using mass spectrometry, is still in its infancy, but the results available to date can be found in Appendix A.4.4.

The subunit mixing procedure was also performed on various mixtures of the wild-type and D398A variant of $GroEL_{14}$. The resulting samples contain $GroEL_{14}$ particles comprised of slow (D398A) and fast (wild-type) ATP hydrolyzing subunits. Since the mixture of subunits is stochastic, the distribution of the various species within a sample can be determined from knowledge of the relative proportion of each variant in the sample.

The samples of mixed subunit $GroEL_{14}$ is labeled with donor fluorophores and pre-steady state kinetic experiments are performed exactly as described in the experiments of Chapter 3.

However, the mixed subunit samples contain subunits that hydrolyze ATP on two very different time scales. Within ten seconds after mixing, more than 98% of wild-type subunits have undergone hydrolysis, compared to less than 1% of the D398A subunits. Thus, on this timescale, the average number of inorganic phosphate released is directly proportional to the mole fraction of wild-type subunits.

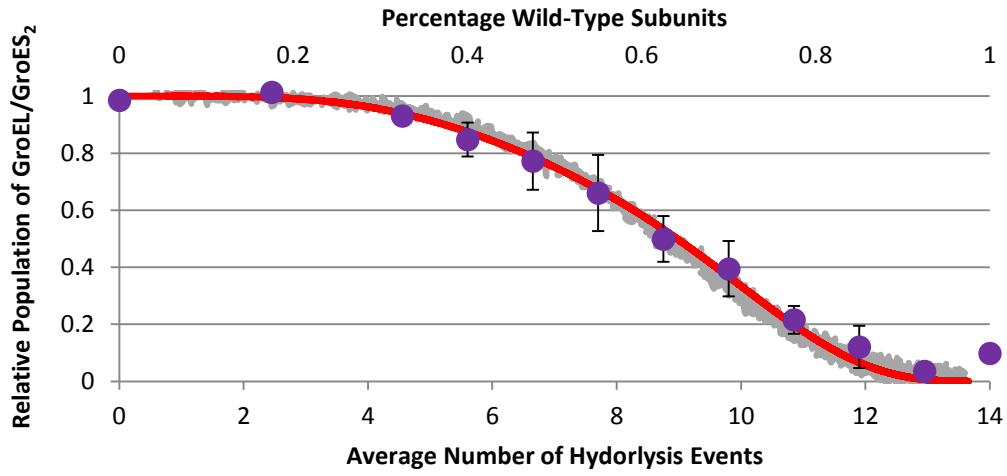
The individual species in the mixed subunit samples are "paused" in a particular nucleotide configuration on the experimental timescale monitored, having undergone only a fraction of the full 14 hydrolysis events. The fluorescence intensity observed after 10 seconds can be converted to the proportion of particles in the sample that remain in their symmetric form, i.e. have not broken symmetry. The subunit mixing experiments thus provide an experimental approach to determining the dependence of symmetric particle population on the average number of observed hydrolysis events that is distinct from the FRET verse Pi experiments described in Section 4.3.5.2.1.

4.3.5.2.3 Comparison of Experimental Approaches

The two experimental data sets that will be used to determine the solution to the breakage of symmetry problem are presented in Figure 4.7.A. The theory developed to address the breakage of symmetry problem quantitates the population of symmetric GroEL/GroES$_2$ complex as a function of the average number of hydrolysis events observed. In the FRET versus Pi experiments, the population of symmetric particles is directly monitored using FRET and the average number of hydrolysis events is directly monitored by measuring the accumulation of inorganic phosphate released. In the subunit mixing experiments, the population of symmetric particles is again directly measured by the average number of hydrolysis events is indirectly measured through the proportion of wild-type subunits within the sample. Despite the experimental differences, both data sets are in good agreement and both are used in the following

Figure 4.7 – Fit of Solution to Breakage of Symmetry Problem to Experimental Data

A



B

$$BoS(X,Y) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

(A) The relative population of symmetric particles is plotted against (lower axis) **the average number of hydrolysis events** observed in a **FRET versus Pi** experiment and (upper axis) the **proportion of wild-type subunits** in a series of **subunit mixing** experiments. The two experiments are in good agreement. Also plotted is the **best fit curve** of Eq 4.14 to the experimental data sets, obtained by optimizing the $P_{ES}(X,Y)$ matrix in the equation. (B) The best fit $P_{ES}(X,Y)$ matrix is presented as the solution to the breakage of symmetry problem.

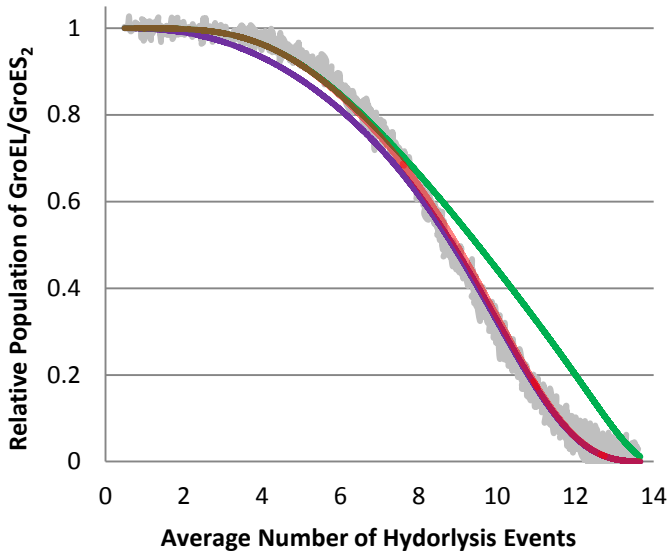
section to fit a solution to the breakage of symmetry problem.

4.3.5.3 Fitting the Solution to the Breakage of Symmetry Model

In Section 4.3.3, it was shown that the solution to the breakage of symmetry problem could be represented as an 8 x 8 matrix of zeros and ones. Assumptions laid out in that section allowed for the reduction of the fitting problem to one of specifying a set of eight cutoff values, $\{w_Y\}_{Y=0}^{7}$. The value of $w_7 = 2$ was determined by single-ring turnover experiments on the D398A variant of GroEL. Specifying the remaining seven cutoff values determines the sought-after solution to the breakage of symmetry problem, Eq 4.8.

Figure 4.8 – Error Analysis of Breakage of Symmetry Solution

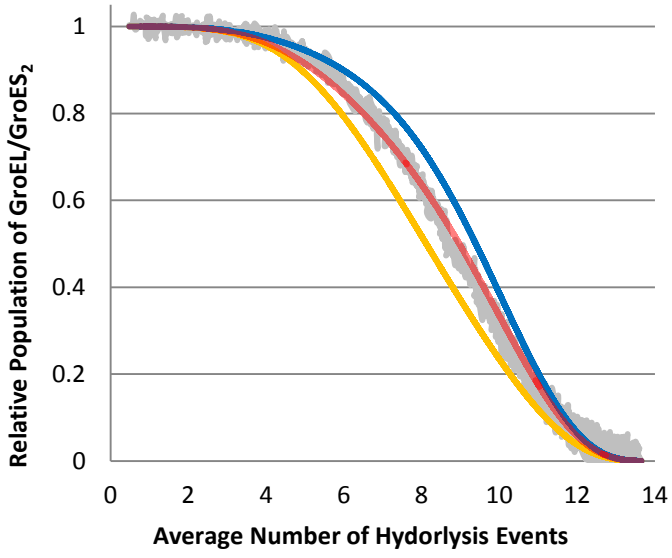
(A) Perturbing the End States: **Matrix A** and **Matrix B**



$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & [1] & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ [1] & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & [0] & 1 \\ 1 & 1 & 1 & 1 & 1 & [0] & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

(B) Perturbing the Middle States: **Matrix C** and **Matrix D**



$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & [1] & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & [1] & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & [0] & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & [0] & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

The following plots are a qualitative error analysis of the **best fit** $\boldsymbol{BoS(X,Y)}$ presented in Figure 4.7.B. The relative population of symmetric particles is plotted against the average number of hydrolysis events observed in the **FRET versus Pi** experiments, performed by Xiang Ye, as well as the solution to Eq 4.14 obtained using the **best fit** $\boldsymbol{BoS(X,Y)}$. (A) Perturbation made by using **Matrix A** and **Matrix B** in place of the **best fit** $\boldsymbol{BoS(X,Y)}$ in Eq 4.14. (B) Perturbation made by using **Matrix C** and **Matrix D** in place of the **best fit** $\boldsymbol{BoS(X,Y)}$ in Eq 4.14. All perturbations produce deviations between the theory and the **FRET versus Pi experimental data**.

The remaining values of $\{w_Y\}_{Y=1}^{6}$ can be fit to the experimental data by minimizing the square of the difference between the two data sets and Eq 4.14. The results are presented in Figure 4.7, along with the final solution to the breakage of symmetry problem, $BoS(X,Y)$, represented in matrix form. Fits to the two experimental data sets, performed individually, led to the same form

of the solution. The final form of the $BoS(X,Y)$ was further supported by fittings that allowed the value of w_7 to float.

4.3.5.4 $BoS(X,Y)$ Error Analysis

In an effort to analyze the error associated with our choices of the cutoff values $\{w_Y\}_{Y=0}^{7}$ that specify the solution, $BoS(X,Y)$, the output of the function in Eq 4.14 was examined using a series of $P_{ES}(X,Y)$ matrices that are closely related to the $BoS(X,Y)$ solution. Each of the four matrices tested have either one more or one less entry than the breakage of symmetry matrix. The results are given in Figure 4.8. The four matrices presented constitute the full set of smallest possible perturbations to the best fit $BoS(X,Y)$ matrix that are consistent with the assumptions stated in Section 4.3.3.

The error analysis performed demonstrates that each of the four $P_{ES}(X,Y)$ matrices tested cause deviations between the theory and the data in a specific isolated region of the curve's domain. For all matrices tested, with the exception of matrix A, the observed deviations are significant, suggesting a high degree of confidence in the solution to the proposed breakage of symmetry problem. Matrix A does increase the discrepancy between experiment and theory, just not to the same degree as the other three matrices tested. Matrix A differs from the best fit $BoS(X,Y)$ in that states $\langle 7,3\rangle$ and $\langle 3,7\rangle$ are no longer stable as a symmetric particles.

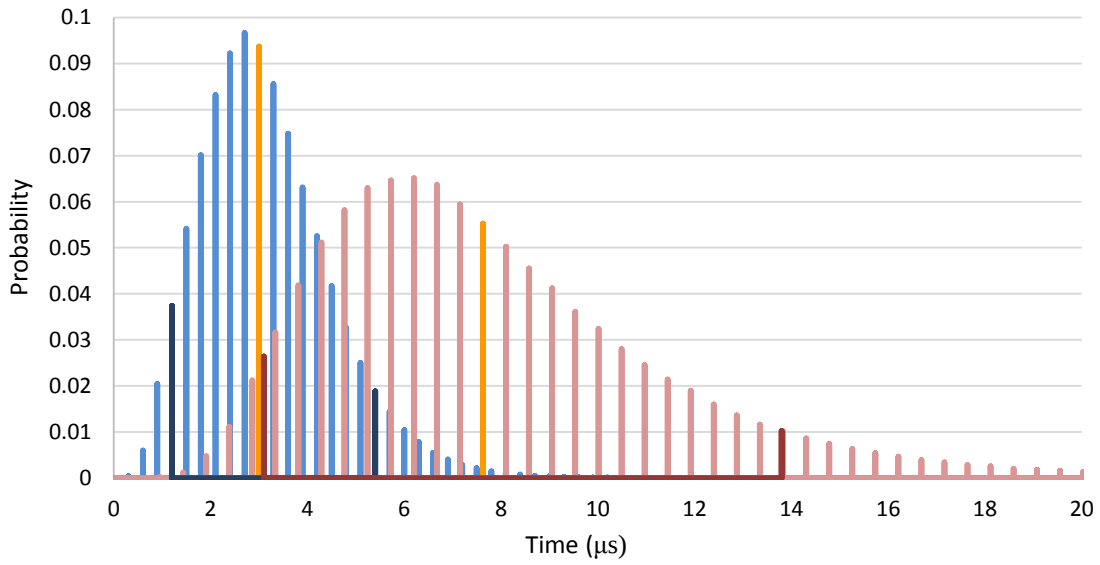
4.3.6 Single Molecule Comparison of Breakage of Symmetry Models

In the previous chapter, single molecule simulations were used to illustrate the differences between the concerted and stochastic hydrolysis model. The analysis was performed before the breakage of symmetry problem was formulated and it was assumed that all seven ATP within a ring must be hydrolyzed before GroES can dissociate. This corresponds to a $P_{ES}(X,Y)$ matrix of all zeros except the bottom row and right-most column, which contain all zeros. The single

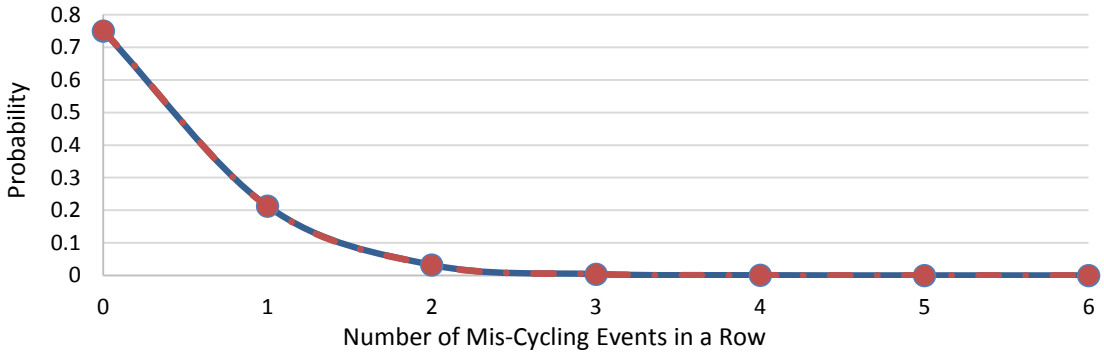
molecule simulations can now be revisited, incorporating the results of the newly proposed solution to the breakage of symmetry problem.

Figure 4.9 – Comparison of Single Molecule Simulation Results for the Different Breakage of Symmetry Models

(A) Distribution of $GroEL_7$/GroES Complex Lifetimes



(B) Number of Mis-Cycling Events



Single molecule simulation comparison of different breakage of symmetry models. Both systems perform hydrolysis of ATP stochastically and have high release micro-constant values, simulating the presence of SP. The difference between the curves is one **requires all seven ATP within a ring be hydrolyzed** before GroES can dissociate and the other utilizes **the solution to the breakage of symmetry model depicted in Figure 4.7.B**. (A) Distribution of complex lifetimes for the two systems along with their **mean** encapsulation times. The 90% confidence range is also indicated on both curves in the corresponding dark colors. The application of **the solution to the breakage of symmetry model in Figure 4.7.B** shifts the decreases the **mean** encapsulation time and narrows the distribution by a factor of 2.5. (B) Despite the change in encapsulation times observed, the number of unproductive mis-cycling events does not increase when the breakage of symmetry model is changed. The curve corresponding to the breakage of symmetry model that **requires all seven ATP within a ring be hydrolyzed** is presented as a dotted line so the other curve can be seen.

Figure 4.9 contains a comparison of the two breakage of symmetry solutions discussed. The addition of the solution to the breakage of symmetry problem to the single molecule simulations shifts the distribution of complex lifetimes towards shorter encapsulation times. Both the mean encapsulation and the width of the 90% confidence interval are decreased by a factor of 2.5. Despite these changes, the rate of mis-cycling is not increased. Thus, the breakage of symmetry solution proposed in this chapter lead to a shorter and more narrowly distributed set of encapsulation times compared to the all-or-none breakage of symmetry model, without increasing the occurrence of encapsulation errors.

4.4 CONCLUSION

The central role that the symmetric GroEL/GroES$_2$ species plays in catalyzing SP folding has only recently been appreciated (12-15). The stochastic mechanism of ATP hydrolysis implies the population of symmetric complexes is actually an ensemble of species with varying nucleotide distributions. This goal of this chapter was to establish a relationship between the nucleotide state of the symmetric complex and the propensity of GroES dissociation. The mechanism of complex symmetry breakage plays a fundamental role in the folding catalysis of the GroEL/GroES system, driving the exchange of GroES and dictating the cycling of SP encapsulation and release.

The first approach to addressing the breakage of symmetry problem was the development of a time-dependent model that postulated a kinetic competition between GroES dissociation and ATP hydrolysis within a ring. Comparison of the symmetric complex decay of the wild-type and D398A variant of GroEL demonstrated that the breakage of symmetry was a time-independent process: nucleotide states either do or do not form stable symmetric complexes. The breakage of symmetry problem can thus be solved by identifying the stable/unstable symmetric nucleotide states.

An additional benefit of the time-independence of the breakage of symmetry problem is the elimination of the need to consider the hydrolysis micro-constant as a parameter. A theoretical relationship was established between the size of the symmetric complex population and the average number of hydrolysis events observed, both experimentally measurable quantities. The theory links the two quantities in a manner that only depends on the initial hydrolysis state of the system and the desired mechanism of symmetry breakage. Two distinct sets of experimental data were fit, both of which produce the same final solution to the breakage of symmetry problem. Error analysis of the solution demonstrates that the fit is quite sensitive even to minor perturbations in the input matrix, increasing with the number of hydrolysis events observed.

The solution to the breakage of symmetry problem obtained indicates that, in the presence of SP, a ring in the symmetric complex never undergoes a full round of hydrolysis before GroES dissociates. This finding runs contrary to popular opinion. The results also demonstrate the importance inter-ring communication in the cycling of the $GroEL_7$/GroES complex. The probability of GroES dissociation from a particular ring within a given symmetric complex is dependent on the ATP occupancy of the adjacent rings.

The analysis of the GroEL/GroES breakage of symmetry problem has led to a curious new species that is an obligatory state along the GroEL/GroES catalytic cycle. When an unstable nucleotide state is reached, GroES is ejected from the symmetric complex, but the newly formed *trans* ring still contains a small number of un-hydrolyzed molecules of ATP. It's well established that the R state of GroEL has a lower affinity for substrate protein (16) and it may be that the function of this species is to help facilitate the release of SP from the central cavity by preventing immediate rebinding.

4.5 REFERENCES

1. Azem, A., M. Kessel and P. Goloubinoff. "Characterization of a Functional

Groel14(Groes7)2 Chaperonin Hetero-Oligomer." *Science* 265, no. 5172 (1994): 653-6.

2. Azem, A., S. Diamant, M. Kessel, C. Weiss and P. Goloubinoff. "The Protein-Folding Activity of Chaperonins Correlates with the Symmetric Groel14(Groes7)2 Heterooligomer." *Proc Natl Acad Sci U S A* 92, no. 26 (1995): 12021-5.

3. Engel, A., M. K. Hayer-Hartl, K. N. Goldie, G. Pfeifer, R. Hegerl, S. Muller, A. C. da Silva, W. Baumeister and F. U. Hartl. "Functional Significance of Symmetrical Versus Asymmetrical Groel-Groes Chaperonin Complexes." *Science* 269, no. 5225 (1995): 832-6.

4. Harris, J. R., A. Pluckthun and R. Zahn. "Transmission Electron Microscopy of Groel, Groes, and the Symmetrical Groel/Es Complex." *J Struct Biol* 112, no. 3 (1994): 216-30.

5. Llorca, O., J. L. Carrascosa and J. M. Valpuesta. "Biochemical Characterization of Symmetric Groel-Groes Complexes. Evidence for a Role in Protein Folding." *J Biol Chem* 271, no. 1 (1996): 68-76.

6. Llorca, O., S. Marco, J. L. Carrascosa and J. M. Valpuesta. "Symmetric Groel-Groes Complexes Can Contain Substrate Simultaneously in Both Groel Rings." *FEBS Lett* 405, no. 2 (1997): 195-9.

7. Lorimer, G. "Protein Folding. Folding with a Two-Stroke Motor." *Nature* 388, no. 6644 (1997): 720-1, 723.

8. Xu, Z. and P. B. Sigler. "Groel/Groes: Structure and Function of a Two-Stroke Folding Machine." *J Struct Biol* 124, no. 2-3 (1998): 129-41.

9. Horowitz, G. H. Lorimer and Ybarra. "GroES in the Asymmetric GroEL/ GroES Complex Exchanges Via an Associative Mechanism." *PNAS* 96, no. 6 (1999): 2682.

10. Grason, J. P., J. S. Gresham and G. H. Lorimer. "Setting the Chaperonin Timer: A Two-Stroke, Two-Speed, Protein Machine." *Proc Natl Acad Sci U S A* 105, no. 45 (2008): 17339-44.

11. Xu, Z., A. L. Horwich and P. B. Sigler. "The Crystal Structure of the Asymmetric Groel-Groes-(Adp)7 Chaperonin Complex." *Nature* 388, no. 6644 (1997): 741-50.

12. Koike-Takeshita, A., M. Yoshida and H. Taguchi. "Revisiting the Groel-Groes Reaction Cycle Via the Symmetric Intermediate Implied by Novel Aspects of the Groel(D398a) Mutant." *J Biol Chem* 283, no. 35 (2008): 23774-81.

13. Sameshima, T., R. Iizuka, T. Ueno and T. Funatsu. "Denatured Proteins Facilitate the Formation of the Football-Shaped Groel-(Groes)2 Complex." *Biochem J* 427, no. 2 (2010): 247-54.

14. Takei, Y., R. Iizuka, T. Ueno and T. Funatsu. "Single-Molecule Observation of Protein Folding in Symmetric Groel-(Groes)2 Complexes." *J Biol Chem* 287, no. 49 (2012): 41118-25.

15. Ye, X. and G. H. Lorimer. "Substrate Protein Switches Groe Chaperonins from

Asymmetric to Symmetric Cycling by Catalyzing Nucleotide Exchange." *Proc Natl Acad Sci U S A* 110, no. 46 (2013): E4289-97.

16. Horowitz, G. H. Lorimer and Ybarra. "Groes in the Asymmetric Groel Groes Complex Exchanges Via an Associative Mechanism." *PNAS* 96, no. 6 (1999): 2682.

17. Mendoza, J. A., B. Demeler and P. M. Horowitz. "Alteration of the Quaternary Structure of Cpn60 Modulates Chaperonin-Assisted Folding. Implications for the Mechanism of Chaperonin Action." *J Biol Chem* 269, no. 4 (1994): 2447-51.

18. Mendoza, J. A., J. L. Martinez and P. M. Horowitz. "Tetradecameric Chaperonin 60 Can Be Assembled in Vitro from Monomers in a Process That Is Atp Independent." *Biochim Biophys Acta* 1247, no. 2 (1995): 209-14.

19. Ybarra, J. and P. M. Horowitz. "Inactive Groel Monomers Can Be Isolated and Reassembled to Functional Tetradecamers That Contain Few Bound Peptides." *J Biol Chem* 270, no. 39 (1995): 22962-7.

20. Ybarra, J. and P. M. Horowitz. "Refolding and Reassembly of Active Chaperonin Groel after Denaturation." *J Biol Chem* 270, no. 38 (1995): 22113-5.

21. Gorovits, B. M., J. W. Seale and P. M. Horowitz. "Residual Structure in Urea-Denatured Chaperonin Groel." *Biochemistry* 34, no. 42 (1995): 13928-33.

Chapter 5

Molecular Dynamics Simulations of GroEL's Allosteric Transitions

5.1 INTRODUCTION

5.1.1 Background

Allostery is the keystone of the GroEL/GroES system, critical not only to its catalytic function, but its ability to dynamically respond to environmental stimuli. The iterative annealing model credits the T → R allosteric transition as being responsible for the forced unfolding of substrate proteins, affording the SP a new opportunity to reach the native state (1-3). The ability of bound SP to stabilize the T state (4), promoting nucleotide exchange and, decreasing the mean cycle time (5,6), allows the GroEL/GroES system to respond to the needs of the cell, entering the active symmetric cycle (Figure 2.13) only when necessary (7-9). This dynamic response feature of system is governed by the R → T allosteric relaxation. Despite a great deal of experimental attention on the T → R transition over the years, very little has been done to investigate the allosteric relaxation, which has largely been assumed to simply be the reverse mechanism of the forward transition.

Numerous studies have established that the T → R allosteric transition is largely concerted, providing the "power-stroke" required to forcibly unfold bound SP and discharge it into the central cavity (10-12). The concerted nature of GroEL's allostery is thought to be a direct consequence of the architecture of the heptameric ring (10). Crystallographic analysis has demonstrated partial overlap of subunits in the direction of the allosteric transition, leading to steric clashes that would prevent a sequential allosteric transition. Other studies have demonstrated the importance of inter and intra-subunit salt bridges in coupling intra-ring domain movements (14). All these experiments demonstrate the concerted domain movements of the T → R transition but the findings do not necessarily naturally extend to other aspects of the catalytic cycle.

This idea of concerted domain movements has haplessly pervaded many other aspects of GroEL/GroES theory. Chapter 3 showed that the widely held belief that ATP hydrolysis within the system is concerted is likely not the case, and a stochastic mechanism is better suited to describe the ATPase behavior of the enzyme. Chapter 2 demonstrated that the ADP release portion of the catalytic cycle is a multi-step process, a finding that presumably implicates the R → T allosteric relaxation as an important and possibly rate limiting step in the absence of SP. A recently published crystal structure of the GroEL/ADP complex was found to be in an asymmetric relaxed state: hard evidence that the movement of subunits within a ring is not necessarily synchronized. It now appears possible that the well-established concerted movement of domains only hold true for the forward allosteric transition, and intra-ring asymmetry plays a fundamental role in the properties of the R → T allosteric relaxation.

5.1.2 Overview

The final chapter of my thesis marks a pivot in my research career from *in vitro* to *in silico* experimentation. The focus of the chapter is a series of closely related molecular dynamics (MD) simulations that aim to investigate the allosteric transitions of GroEL. These MD simulations monitor a coarse grained model of GroEL's atomic structure as it transitions between the T and R allosteric states. The results of the simulations allow us to characterize the transition in terms of key domain movements and the time scales that separate them. In the subsequent analysis, particular attention is paid to ring symmetry and how it differs between the forward T → R transition and reverse R → T allosteric relaxation. These simulations provide insight into the key and distinctive roles that inter-subunit contacts play in the driving the forward and reverse allosteric transitions.

The first portion of the chapter is dedicated to detailing the simulation method applied to modeling the allosteric transitions in GroEL, which closely follows previously published work by

Hyeon et. al. (15). The self-organized polymer (SOP) model is used to coarse grain the structure of GroEL. The Langevin equation of motion is used to describe the dynamics of the system, without the explicit representation of solvent molecules. The Hamiltonian used in the equation of motion contains a number of allosteric state dependent parameters, and switching these parameters over the course of the simulation drives the transition between states. The method has been demonstrated to produce realistic simulations of domain movements.

Each MD simulation yields an abundance of data that has to be filtered in a rational manner so that the information of interest can be easily extracted. A computer program was written that reads each simulation output and organizes the coordinates into the various domains known for GroEL. Each subunit is then fit with its own coordinate system, allowing for easy comparison of subunits within a ring. Finally, a set of simulation results for a particular domain is then analyzed using singular value decomposition (SVD) to determine a reduced set of variables that best describes the motion of that domain (16).

Application of SVD to the simulation results affords us an opportunity to quantitatively evaluate the assumption that GroEL's allosteric transition is comprised of a series of rigid domain movements. Assessment of the assumption of rigid body motion also allows us to identify flexible regions within the domains. The set of domain measures, obtained through SVD, allows us to compare the movement of subunits within a ring, and across multiple simulations. The ensemble average domain motions are analyzed to further characterize the transitions.

In addition to characterizing the movement of domains within a subunit, the relative motion of subunits within a ring is also considered. Simulation results are evaluated with regard to the symmetry of intra-ring domain movements. Asymmetry in the forward and reverse allosteric transitions are compared, revealing distinctive differences between the two. Asymmetry within the reverse R → T transition is then further analyzed, revealing dynamic negative inter-subunit

cooperativity, which may prove to be the cause of the multi-step ADP release processes observed in Chapter 2.

5.2 METHODS

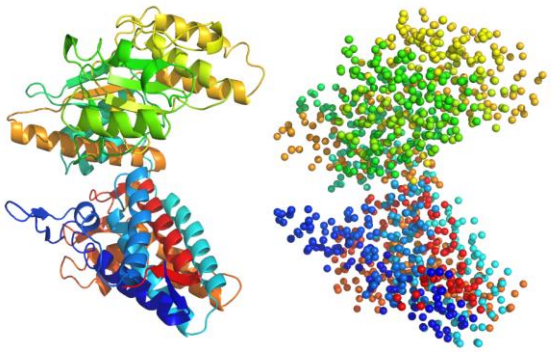
5.2.1 Simulation Details

5.2.1.1 Choice of Crystal Structure

The simulations performed sketched out plausible paths taken by subunits as they complete the large scale movements that characterize the allosteric transition of GroEL. The simulations are structure based, utilizing previously solved crystal structures of the T and R states as references. The beginning and endpoint of each simulation, the T and R states, are predetermined by the crystal structures and input into the MD program. The focus of the simulations of this type is not where the transitions lead, but the path they take to get there. Numerous crystal structures exist for the T state that do not differ significantly. The 1KPO.pdb structure was chosen because it contained coordinates for both rings (17).

More controversial is the choice of the R state structure. Several structures exist that differ quite significantly (17-20). The structure chosen to represent the R state was selected on the basis that it was the most distinct from the T state, the so-called ROpen structure. The ROpen subunit has been observed within a symmetric ATP-bound ring using symmetry imposed cryo-EM at a resolution of 8.5 Å (19) and within an asymmetric ADP-bound ring using x-ray crystallography at a resolution of 2.7 Å (20). Despite the controversy over whether or not the symmetric ROpen structure exists, it was nonetheless chosen because it best suited the goals of the experiment. Since one of the goals of the study is to investigate the role of asymmetry in the cycle, the use of an asymmetric end state such as the ADP bound structure would be counterproductive. Also, looking closely at the asymmetric ADP bound structure, it appears that all other subunits contain

Figure 5.1 – Illustration of the Self-Organized Polymer Model of a GroEL Subunit



A GroEL subunit before (left) and after (right) the SOP course graining procedure, outlined in Section 5.2.1.2, was applied. The residues are colored by sequence for easier comparison.

some distinct mixture of features held by the T state and the ROpen state. Thus, the ROpen structure was chosen on the basis that it represents the "most R-like" structure, containing the maximal amount of features that differentiate a T subunit from an R subunit.

5.2.1.2 Self-Organized Polymer Model

In order to run a molecular dynamics simulation on GroEL, the structure has to be coarse grained due to its size. The self-organized polymer model (SOP) is used to represent the structure of GroEL. The SOP model has been shown to realistically simulate not only the mechanical motions in proteins (21-23) but also map their free energy landscapes (15, 24-25). In the SOP model, the structure of each GroEL subunit is represented by a chain of "beads". Each residue in the protein has two corresponding beads in the simulation. The first bead represents the backbone atoms of the residue and strongly interacts with the backbone beads of its neighboring residues. Its position is determined by the C_α position of the corresponding residue in the state's crystal structure. The second bead represents the side chain of the residue and has a soft sphere radius corresponding to the size of the particular residue's side chain. Its position is given as the center of mass of the side chain atoms, as determined by the crystal structure. An illustration of the SOP model of a GroEL subunit is given in Figure 5.1.

5.2.1.3 Equation of Motion

In the following simulations, the motion of the SOP chains is governed by Langevin dynamics in the over-damped limit (26). Newtonian dynamics is well suited to describe the behavior of

molecules in a vacuum but proteins move in solution, constantly colliding with solvent molecules which damp their motion. Langevin dynamics has a term that accounts for damping of motion by solvent molecules, a velocity-dependent friction term, as well as a term that accounts for a contribution to the motion due to thermal energy from solvent collisions. In the over-damped limit, the acceleration-dependent term is negligible and can be dropped from the equation of motion. The integrator used in the MD simulations presented in this chapter is given in Eq 5.1. The movement of a bead in some small time step, Δt, corresponds to the quantity $\vec{r}(t + \Delta t)$, which is dependent on its position, $\vec{r}(t)$, and the forces acting on it from all other beads, $\vec{F}(t|\, X)$. The force depends on the current allosteric state of the system, indicated by the parameter X. The parameter ζ is the friction coefficient of the over-damped system and is given a value of 17 ps^{-1}. The final term in the equation of motion corresponds to the random thermal motion imparted on the protein from the solvent. The function $\Gamma(t)$is a randomly generated white noise spectrum, which leads to a unique set of results for each simulation performed.

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \frac{\Delta t}{\zeta}\vec{F}(t|\, X) + \sqrt{2\,\frac{\Delta t}{\zeta}\,k_B T}\,\Gamma(t) \qquad \text{Eq 5.1}$$

5.2.1.4 Hamiltonian

The force acting on a bead is given as the negative of the gradient of the state dependent Hamiltonian, $\vec{F}(t|\, X) = -\nabla H(\{\vec{r}_i\}|\, X)$. The Hamiltonian used is presented in Eq 5.2 (10). The first term is the finite extendable non-linear energy (FENE) term that represents the covalent bond between adjacent backbone beads and, between side chain and backbone beads. The third term is the soft sphere repulsion term that prevents beads from coming into too close of contact as well as chain crossing. The value of $\epsilon_{L\,i,j}$ is constant for all residues and σ_{ij}, is a term determined by the soft sphere radius of the two amino acids. The final term is the Coulombic energy between charged residues and includes a Debye-Huckel shielding factor to account for the shielding due to

solvent molecules. More specific details about the exact values of all constant parameters used throughout these MD simulations can be found in Appendix A.5.1.

$$H\left(r_{ij}|\,X\right) = \sum_{i,j\ Bonded} \frac{-k}{2}\, R_o^2\, \log\left(1 - \frac{\left(r_{ij} - r_{ij}^o(X)\right)^2}{R_o^2}\right)$$

$$+ \sum_{i,j\ Native\ Contact} \epsilon_{N\,i,j} \left(\left(\frac{r_{ij}^o(X)}{r_{ij}}\right)^{12} - 2\left(\frac{r_{ij}^o(X)}{r_{ij}}\right)^{6}\right)\Delta_{ij}(X)$$

$$+ \sum_{i,j\ Native\ Contact} \epsilon_{L\,i,j} \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \left(1 - \Delta_{ij}(X)\right)$$

$$+ \sum_{i,j\ Charged} \frac{q_i q_j}{4\,\varepsilon\, r_{ij}}\, e^{-\kappa\, r_{ij}} \qquad \text{Eq 5.2}$$

The parameter $\Delta_{ij}(X)$ in the second and third terms takes on the value of one if residues *i* and *j* are within 8 Å in state X and a value of zero otherwise. The second term of the Hamiltonian is an attractive potential between native contact residues. It takes the functional form of the Lennard-Jones potential with the distance parameter $r_{ij}^o(X)$ corresponding to the distance between residues i, j in state X. The use of an attractive potential of this form requires an *a posteriori* justification of the nature of protein folding energetics. The folding potential energy landscape of most proteins is relatively smooth and adequately maps to a single coordinate: the number of native contacts. By extension, the energetics of the structure of an allosteric state can then defined by its specific set of native contacts. The energy factor $\epsilon_{N\,i,j}$ is determined from the knowledge-based Betancourt-Thirumalai potential (27). The Betancourt-Thirumalai potential is dependent on the identity of residues i and j, assuming a different value for each unique amino acids pair. More detailed information about how the value of $\epsilon_{N\,i,j}$ is calculated can be found in Appendix A.5.2.

The Hamiltonian for each allosteric state can be found by extracting the structure dependent terms from the PDB file of that state. If one begins with an SOP model of a particular allosteric state, say the X state, and a simulation is run using the Hamiltonian in the equation of motion that corresponds to the X state, ie $H\left(r_{ij}|\, X\right)$, the simulated motion would correspond only to thermally driven fluctuations in the structure. Each simulation is begun by this sort of thermally equilibration. The allosteric transition is induced by changing the structure dependent terms of the Hamiltonian, $r_{ij}^{o}(X)$, from one allosteric state to another. The change in the underlying potential function forces the SOP modeled GroEL to search for its new minimum potential well. The motion of the SOP is approximately the path of least resistance through the altered potential energy landscape.

Consider a simulation of the $X \rightarrow Y$ transition. The starting structure used would be the SOP model derived from the PDB of the X state structure. Each simulation consists of three parts and the Hamiltonian takes on a different form in each part. The first corresponds to the thermal equilibration mentioned above, where the Hamiltonian takes the form $H\left(r_{ij}|\, X\right)$. After a sufficient equilibration period, the Hamiltonian is switched to the transition form $H\left(r_{ij}|\, T \rightarrow R\right)$. Note that the only parameters in the Hamiltonian that change will be the ones corresponding to the native contact distance, $r_{ij}^{o}(X)$. The transition portion of the simulation consists of N steps and the structure dependent parameter of the nth step is given by Eq 5.3. The value of N used in the simulations presented was 100,000. Additionally, the time step is reduced 10-fold during this portion of the simulation to prevent unstable motion. The Hamiltonian of the final portion of the simulation is simply given by $H\left(r_{ij}|\, Y\right)$. During this phase, the SOP locates its new minimum free energy well at state Y.

$$r_{ij}^{o}(T \rightarrow R) = \left(1 - \frac{n}{N}\right) r_{ij}^{o}(T) + \frac{n}{N} r_{ij}^{o}(R) \qquad \text{Eq 5.3}$$

5.2.1.5 Time-Dependent Simulation Parameters

As is always done in MD studies, the form of the integrator used, Eq 5.1, progresses in time through discrete time steps, Δt. The size of the time step is very important in MD studies. One does not want to choose one too large, or else the beads may begin to display unstable motion, characterized by an unreasonably large change in position for a given time step, eventually resulting in the ejection of atoms from the system. One the other hand, the smaller the choice of Δt, the more data sets that one has to generate to observe the same time period. The time step chosen for the simulations performed was 17.5 ps, and decreased to 1.75ps during the transition portion of the simulation. The coordinates of the beads were output every 10,000 time-steps, 175ns, and a total of $3x10^7$ time steps were performed, yielding ~500μs of simulated protein movement..

5.2.1.6 Control Simulations

Two simulations were run as controls for the T and R allosteric states. These simulations were performed exactly as described above, but with the same structures used for the initial and final states. That is, the simulation corresponds to an $X \rightarrow X$ transition. These simulations are essentially just large time-scale thermal equilibration profiles of the two states. The control simulations are used in the subsequent analysis to form a set of standard distance measure values that correspond to each transition state.

5.2.2 Analysis Details

5.2.2.1 Domain Measures

Allosteric transitions are generally characterized by a large scale reorganization of residues. The goal of this work is to characterize these transitions in a concise and cogent manner. The MD simulations performed generate numerous “simulation outputs”, for each trajectory, made up of

the xyz coordinates of all beads in the SOP model. An allosteric transition is most accurately characterized by the displacement of each atom from its starting position. However, the large number of atoms in a protein structure makes this wealth of information too vast to provide insight. In order to make the information more manageable, the first step is to pre-process the data, condensing the results into a smaller set of measures that fully characterize the observed motions.

In certain special circumstances, the motion of a group of atoms can be completely characterized by a set of six variables in Eq 5.4, corresponding to the displacement of the center of the group, $\vec{r} = \{r_i\}_{i=1}^{3}$, and the rotation about each axis, $\vec{\theta}\{\theta_i\}_{i=1}^{3}$. If a set of coordinates describe the position and orientation of a body, these six variables completely describe the movement of that object only when the object is rigidly transformed, i.e. there is no distortion of the body during the transformation. This special case is a so-called rigid body transformation.

$$R\Theta = \{r_x, r_y, r_z, \theta_x, \theta_y, \theta_z\} \qquad \text{Eq 5.4}$$

Application of the rigid body translation assumption to the domain movements that make up an allosteric transition would allow each simulation trajectory to be reduced to a set of six time-dependent domain movement measures, $R\Theta(t)$. However, proteins are not rigid bodies: thermal motion is constantly driving the protein structure through a series of conformational changes. Therefore, the rigid body assumption can only be applied in special cases. Fortunately, singular value decomposition (SVD) can be used to qualitatively assess the validity of the rigid body translation assumption and generate the six domain measures in Eq 5.4 (16).

Singular value decomposition is a method of matrix factorization of the form $SVD(M) = UZV^T$ where U and V are orthogonal matrices and Z is a diagonal matrix (33). Considering a matrix M as a coordinate transformation, which can both sheer (stretch) and rotate simultaneously. SVD can be thought of as decomposing M's operation into a series of events: an initial rotation,

followed by a distortion, and then a final rotation. The fact that U and V are orthogonal means that they preserve the length of the vectors they operate on, qualifying them as rotation matrices. Then the matrix $\mathcal{Z}$ is solely responsible for the distortion of vector norms observed in the transformation by matrix M. After singular value decomposition of M, the matrices U and V are then re-combined without the rescaling component $\mathcal{Z}$, to generate the rotation matrix, $\hat{R}$.

Let $\{\vec{p}_i\}_{i=1}^N$ be the positions of a subset of residues in state X and $\{\vec{q}_i\}_{i=1}^N$ be the positions of that same set of residues in state Y. Let T: $\vec{p} \rightarrow \vec{q}$ be the transformation that maps the set $\{\vec{p}_i\}$ to $\{\vec{q}_i\}$. The transformation T is a rigid body transformation if and only if it has the form $T = (\hat{R} + \vec{d})$ where $\hat{R}$ is a 3 x 3 rotation matrix and $\vec{d}$ is a three dimensional translation vector. Singular value decomposition can be used to find the components of the transformation T, that minimize the quantity in Eq 5.5; ie the least squares best fit rigid body transformation.

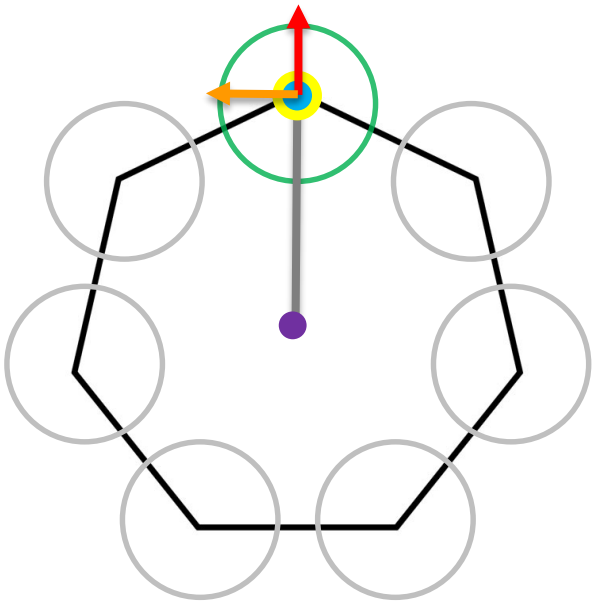
$$Error = \sum_{i=1}^{N} \left\| \hat{R}\vec{p}_i + \vec{d} - \vec{q}_i \right\|^2 \qquad \text{Eq 5.5}$$

This technique is used extensively in the analysis of the simulation results to find the best fit rigid body transformation between an initial set of coordinates, obtained from the average of the control T state simulation, and the set periodically obtained throughout the course of the allosteric transition, i.e. from each simulation output. The best fit transformation $T = (\hat{R} + \vec{d})$ is then used to generate the set of coordinates that arise from a rigid body transformation of the initial set. The theoretically generated coordinates can be compared to the *in silico* experimentally obtained coordinates to determine the validity of the rigid body transformation.

5.2.2.2 Internal Coordinate System

In order to compare the motion of subunits within a ring, measures of domain motion are made relative to a coordinate system that is uniquely defined for each subunit. That is, each subunit is

Figure 5.2 – Cartoon Illustration of the Internal Subunit Coordinate System



Each subunit is fit with its own internal coordinate system so domain movements within a ring can be directly compared. The selected **subunit** is first fit with an **origin** at the center of mass of its equatorial domain (inside yellow circle). The **z-axis**, also called the **normal** direction, is defined normal to the equatorial plate, and points out of the page. The **x-axis**, also called the **radial** direction, is take as the direction from the **seven-fold axis of symmetry** to the **origin**. Lastly, the **y-axis**, or **tangential** direction, is defined by the cross product between the other two directions.

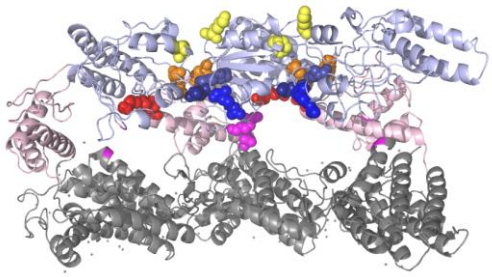
fit with an internal coordinate system at each simulation output and the coordinates of its SOP beads are redefined relative to that coordinate system. The fitting of this coordinate system is based on the presumption that the equatorial domains within a ring are relatively immobile throughout the course of a transition. This has been supported by a large number of structural studies performed on GroEL (20, 28-30). The origin of the internal coordinate system is the center of mass of the equatorial domain of each subunit. The z-axis, or normal direction, is defined as normal to the plane that separates the two rings (the equatorial plate). The x-axis, or radial direction, is defined as running from the ring's seven-fold axis of symmetry to the origin of the subunit. The y-axis, or tangential direction, is defined as the cross product between the z and x axis unit vectors, and points tangentially to the heptameric ring. Using this coordinate system allows us to describe the average motion of the subunit's domains, and easily calculate the variance between the domains within a ring. An illustration of the internal coordinate system of a particular subunit can be found in Figure 5.2.

5.2.2.3 Salt Bridges

The allosteric states of GroEL, particularly the T state, are stabilized by several intra and inter-

Table 5.1 – List of Salt Bridge Pairs Monitored During Analysis

Intra-Subunit					
Asp	83	Equatorial	Lys	327	Apical
Inter-Subunit					
Glu	**386**	**Apical**	**Arg**	**197**	**Inter**
Glu	**216**	**Apical**	**Lys**	**226**	**Apical**
Glu	**232**	**Apical**	**Lys**	**242**	**Apical**
Glu	**257**	**Apical**	**Arg**	**268**	**Apical**
Glu	**257**	**Apical**	**Lys**	**272**	**Apical**



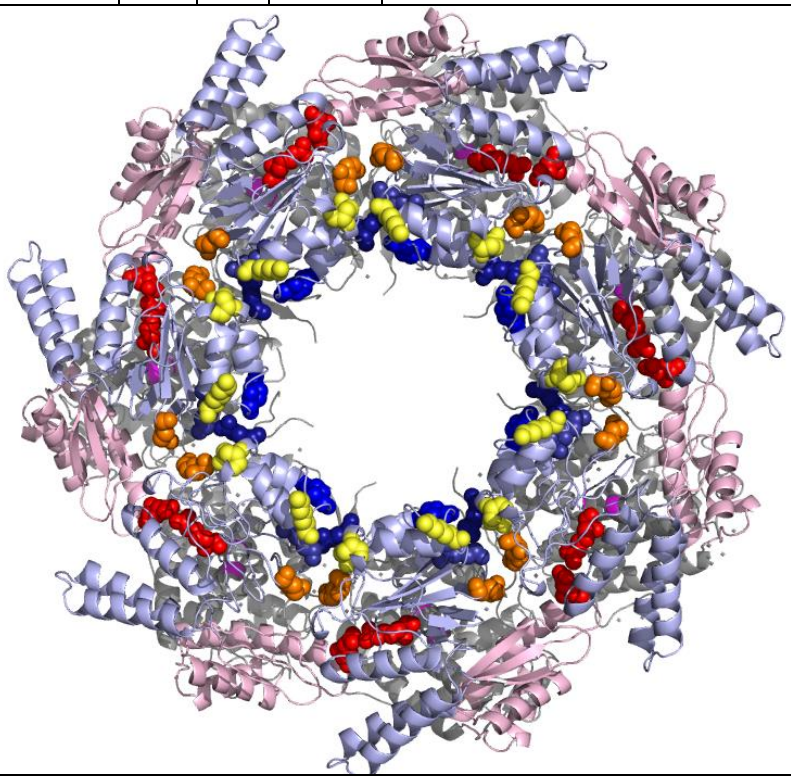


Image of salt bridge pairs followed throughout the MD simulations of GroEL allostery. The **equatorial domain** is presented in grey, the **intermediate domain** in light pink and the **apical domain** in light blue. The logic behind their selection is laid out in Section 5.2.2.3. The only intra-subunit salt bridge is the **83/327** pair. The inter-subunit salt bridges are **386/197**, **216/226**, **232/242**, **257/268** and **257/272**.

subunit salt bridges (14, 20, 28-32). In addition to domain movements, the allosteric transition can be characterized by changes in distance between interacting charged residues. Salt bridges that may be important to the dynamics of the protein were determined by analysis of the crystal structures of the T and R state. The structures were screened for pairs of oppositely charged residues that had a C_α - C_α distance of within 10 angstroms in either allosteric state. This is rather large for a salt bridge, but the measurement was made between α-carbons and not charged atoms within side chain residues. The subset of those pairs whose C_α - C_α distance differed by greater

than 2Å between the two allosteric states were then flagged to be monitored throughout the course of a test MD simulation.

Of the 39 pairs of residues chosen as salt bridges that were potentially key to the dynamics of the allosteric transition, only 6 exhibited a significant change in distance throughout the course of a test simulation. Significance was defined as having a range of distances sampled that is greater than twice the standard deviation of the distances sampled. The only significant intra-subunit salt bridge was D83/K327. This is a well-known salt bridge that stabilizes the T state by linking the apical and equatorial domains. The remaining five salt bridges monitored were located at the inter-subunit domain and are all listed in Table 5.1. It should be noted that, although only the six residue pairs presented were analyzed, the analysis does not claim to identify them as the only six relevant salt bridges in the GroEL system. Distances between other relevant salt bridges may simply be too noisy to analyze accurately in this context.

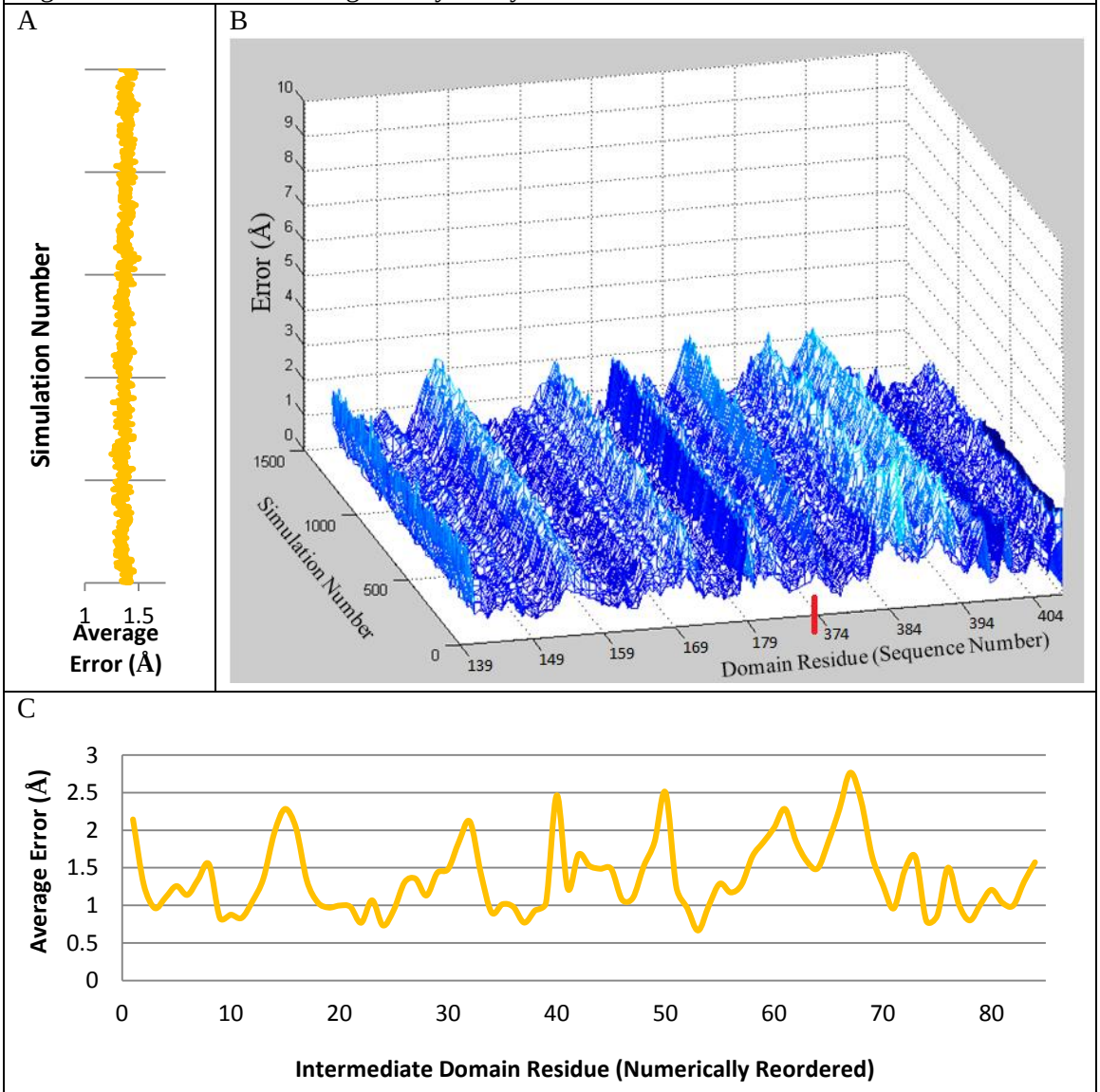
5.3 ANALYSIS

5.3.1 Evaluation of Rigid Motion Assumption

The structure of GroEL is often described in terms of three distinct domains: the equatorial, intermediate and apical domains. Singular value decomposition analysis was performed, as described above, on the subset of residues corresponding to these domains. SVD analysis of a simulation trajectory produces the set of six time-dependent parameters, $\{\vec{r}(t), \theta(t), \psi(t), \varphi(t)\}$, that completely describe the rigid body motion of the domain. These parameters can be plugged back into Eq 5.6 and used to generate a set of coordinates that correspond to the rigid body translation of the initial set of residues.

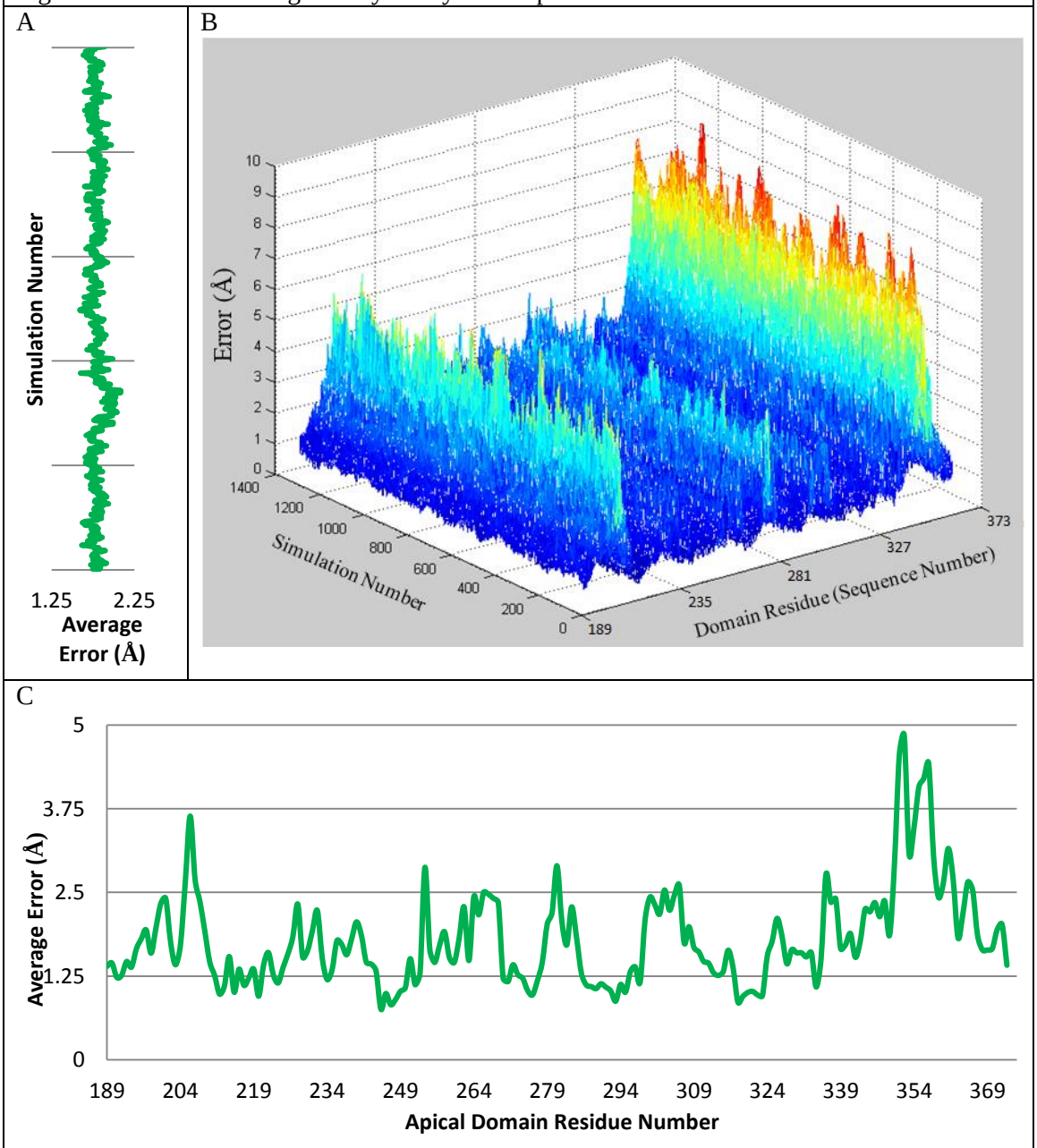
The distance between the coordinates of the rigid body generated by the transformation Eq 5.6, $\{\vec{Q}_i\}$, and the in silico experimentally determined set of residue coordinates, $\{\vec{q}_i\}$, can be

Figure 5.3 – Results of the Rigid Body Analysis of Intermediate Domain Movement



A summary of the results of a quantitative analysis of the assumption rigid body movement of the intermediate domain. The intermediate domain is defined as being composed of residues 139-188 and 375-407. (B) Surface plot of the calculated error of applying the rigid body assumption, as measured by Eq 5.7, to the indicated intermediate domain residue at the given point in a randomly selected simulation of the T → R transition. The intermediate domain is split in sequence space and the break is indicated by a red bar. (A) The surface plot results averaged across simulation number. (C) The surface plot results averaged across residue number. The residue numbers are reordered by one-to-one mapping to the set of natural numbers. Analogous plots for the R → T transition are not given because the results are nearly identical.

Figure 5.4 – Results of Rigid Body Analysis of Apical Domain Movement



A summary of the results of a quantitative analysis of the assumption rigid body movement of the apical domain. The intermediate domain is defined as being composed of residues 189-374. (B) Surface plot of the calculated error of applying the rigid body assumption, as measured by Eq 5.7, to the indicated apical domain residue at the given point in a randomly selected simulation of the T → R transition. (A) The surface plot results averaged across simulation number. (C) The surface plot results averaged across residue number. Analogous plots for the R → T transition are not given because the results are nearly identical.

calculated in order to evaluate the validity of the rigid body assumption, $E_i(t)$, Eq 5.7. The calculated rigid body error is a function of both time, t, and the residue number, i. A surface plot is given for the value of $E_i(t)$ for the residues of the intermediate and apical domains, Figure 5.3 and 5.4, respectively. Both domains show no dependence of $E_i(t)$ on t, and areas of high deviation are isolated in sequence space.

$$\vec{Q}_i(t) = \hat{R}\vec{p}_i(t) + \vec{d} \qquad \text{Eq 5.6}$$

$$E_i(t) = \left\|\vec{Q}_i(t) - \vec{q}_i(t)\right\| \qquad \text{Eq 5.7}$$

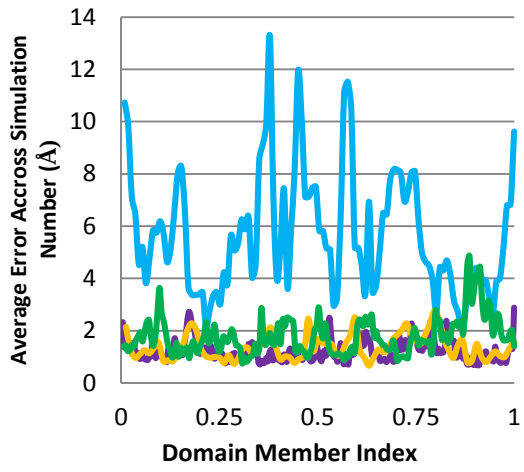
In order to illustrate the validity of the applied analysis technique, a set of 111 residue numbers were randomly generated and the rigid body assumption was applied to residues in the set. The values of $E_i(t)$ are given in Figure 5.5 for this randomly generated set of residues, as well as the three domains of GroEL. The deviation from the rigid body approximation was much higher for the set of random residues. More importantly, the random set shows a clear dependence of $E_i(t)$ on t, increasing throughout the course of the simulation.

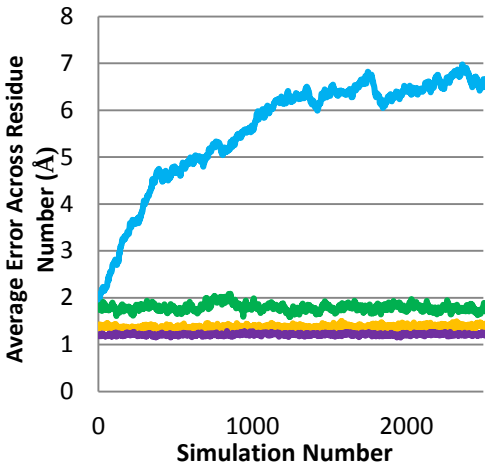
The value of $E_i(t)$ serves as a measure of the deviation of the position of residue i from that predicted by the motion of a rigid body. The isolation of high error points in sequence space, and lack of dependence on simulation progression, indicate that the points of large deviation in the apical and intermediate domains are due to the relative flexibility of the region. The flexible regions can be visualized by coloring the structure of the domains to match the color map in Figure 5.6. For the intermediate domain, most of the deviation is isolated in a loop at the end of helix G. For the apical domain, one region of high deviation is isolated in similar position, at the end of helix J. However, the largest deviation is seen by the hairpin connecting helices K and L. This is likely due to the relative paucity of native contacts of this region in either allosteric state,

Figure 5.5 – Rigid Body Analysis on Random Set of Residues

(A) Deviation from Rigid Body Transformation for each Domain Member

(B) Deviation from Rigid Transformation as a Function of Simulation Progress



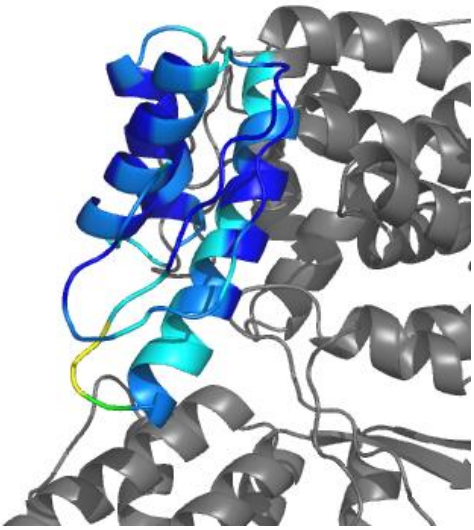


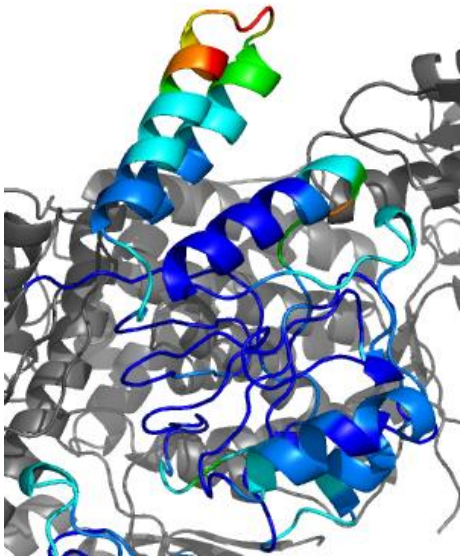
Comparison of the results of applying the rigid body translation assumption to the **equatorial domain** (residues 2-138 and 408-525), the **intermediate domain** (residues 139-188 and 375-407), the **apical domain** (residues 189-407) and, a **randomly generated subset of sequences space** consisting of 111 residues from each domain. The measure of error is calculated according to Eq 5.7. (A) The error is averaged over the various simulation outputs for a randomly selected trajectory. Error values are plotted against a domain member index number, where the sequence number is sequentially mapped to a value between zero and one. (B) Error values are plotted against simulation output number. The increase seen for the **randomly generated subset of sequence space** is what would be expected for a subset of atoms exhibiting non-rigid motion.

Figure 5.6 – Domain Regions with Largest Deviation from Rigid Body Transformation

(A) Intermediate Domain

(B) Apical Domain





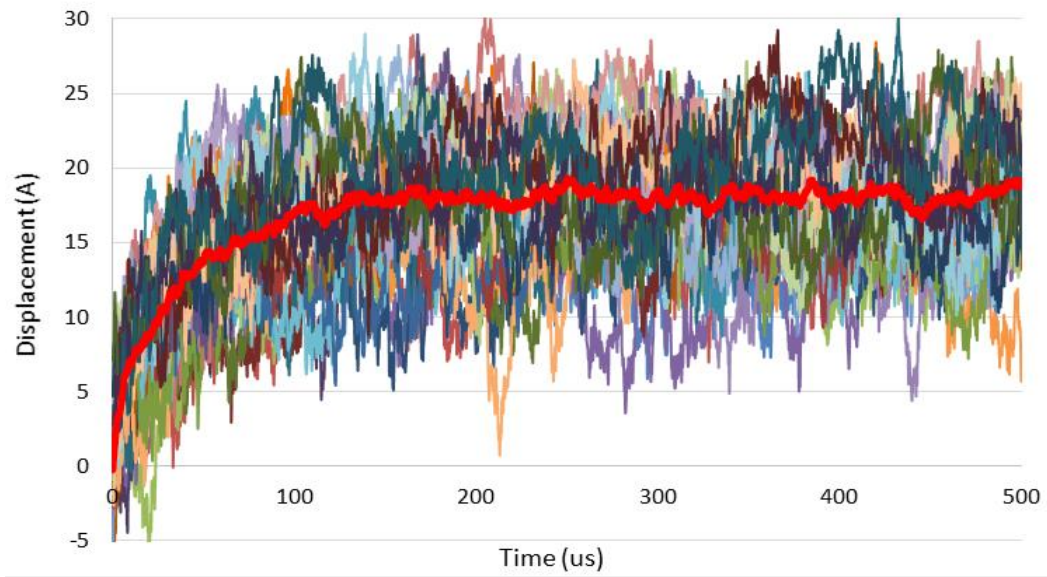
Heat maps of the indicated domain structure, illustrating the error in the rigid body assumption. Color scheme is the same used in Figure 5.4.B, with blue indicating **low error** and red indicating the regions of **high error**. The view is looking directly down the subunit's z axis. The apical domain in panel (A) is removed for clarity. The maps show that, for the most part, the deviation is a result of movement of flexible loop regions in the domain. The exception being observed in panel (B) where the tips of helices K and L also show large deviations.

making it prone to a large amount of thermal motion. In order to ensure their accuracy, these helices were left out during calculations of the domain movement measures reported for the apical domain.

5.3.2 Average Domain Motion

The rigid body analysis performed on the intermediate and apical domains of GroEL have allowed us to conclude that the movement of these domains is well characterized by a rigid body transformation. The error associated with the assumption was shown to be isolated to particular residues and independent of the simulation progression, reflecting the above-average flexibility of these regions. This justifies the use of two simplified sets of domain movement measures to completely characterize GroEL's allosteric transition. The procedure was repeated over 5 distinct simulation trajectories for the T → R transition and 15 trajectories for the R → T allosteric relaxation. Figure 5.7 is an example of the data set produced by the analysis. The plot contains the 35 trajectories simulated of the extension of the intermediate domain radially outward.

Figure 5.7 – Results of the Allosteric Transition T → R Simulation:
Radial expansion of the apical domain



The plot above contains the 35 data sets obtained from performing the SVD analysis described in the test on the structures periodically recorded throughout the simulation. The **average** of all trajectories and can be used to determine the characteristic time of the transition.

5.3.2.1 T → R Allosteric Transition

The results of all simulations can be combined to yield the average domain movements that characterize the allosteric transition. Figure 5.8 gives a plot of the average motions that characterize the apical and intermediate domain movements during the T → R transition. If enough simulations have been averaged, the resulting average motion curves are smooth and well characterized as the sum of exponentials.

In addition to giving the magnitudes of various transitions, the plots contain information on the time scales of the motion. The curves can be fit with exponential functions and the rate terms can be extracted and compared. Table 5.2 contains the rate constants fitted to the average displacement measures of each domain. The number of exponential terms used were determined as the minimum number required to obtain random residuals of the fit. Displacement in the y direction is excluded for the apical domain calculations because the movement averaged around zero throughout the simulation trajectory, indicating a lack of directed motion.

The exponential fitting identifies three distinct time scales of motion. The intermediate domain begins the transition first, first rotating about the y and, most noticeably, z axis, Figure 5.9. The initial rotation of the intermediate domain breaks the inter-subunit contacts with the adjacent subunit, Figure 5.10.B. The newly freed apical domain is then free to move, causing an opening of the apical-intermediate-equitorial angle. This movement breaks the remaining inter-subunit contacts.

Figure 5.10 shows simulation stills at the three different stages of the transition. Note the inter-subunit apical domain contacts before and after the initial fast phase movements (Figure 5.10.A and 5.10.B). After the initial domain movements that separate the subunits, there is a final slow extension of the domains in the direction radially outward and normal to the equatorial plane. This phase is also characterized by slow and pronounced rotations in the apical domain.

Figure 5.8 – Average Domain Movements of T → R Allosteric Transition

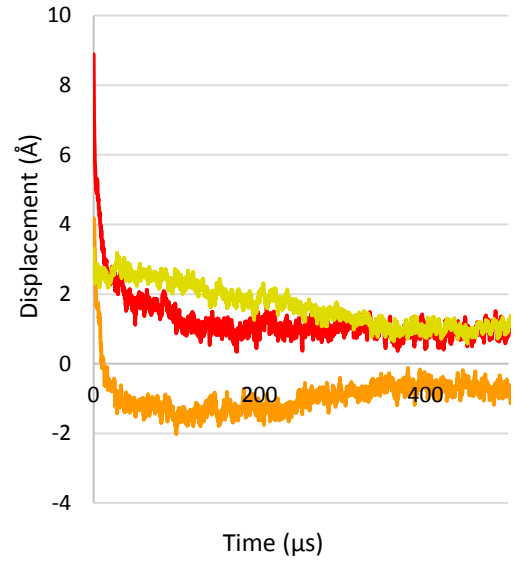
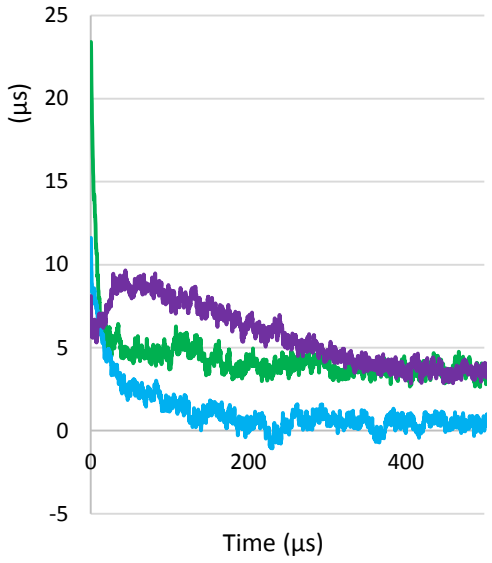
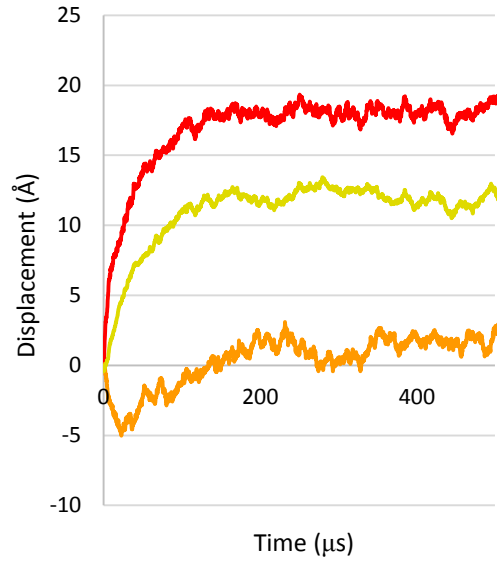
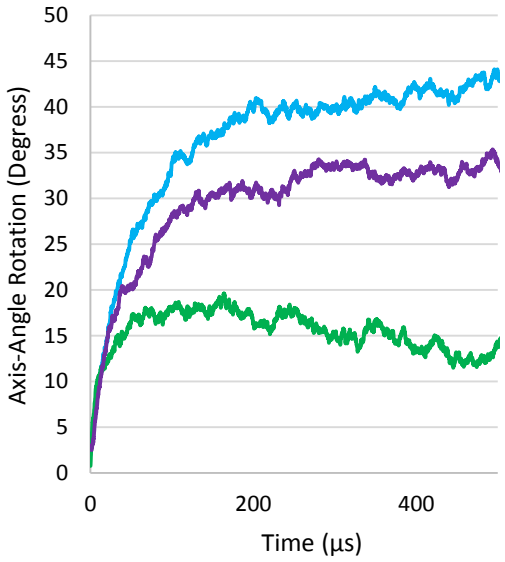
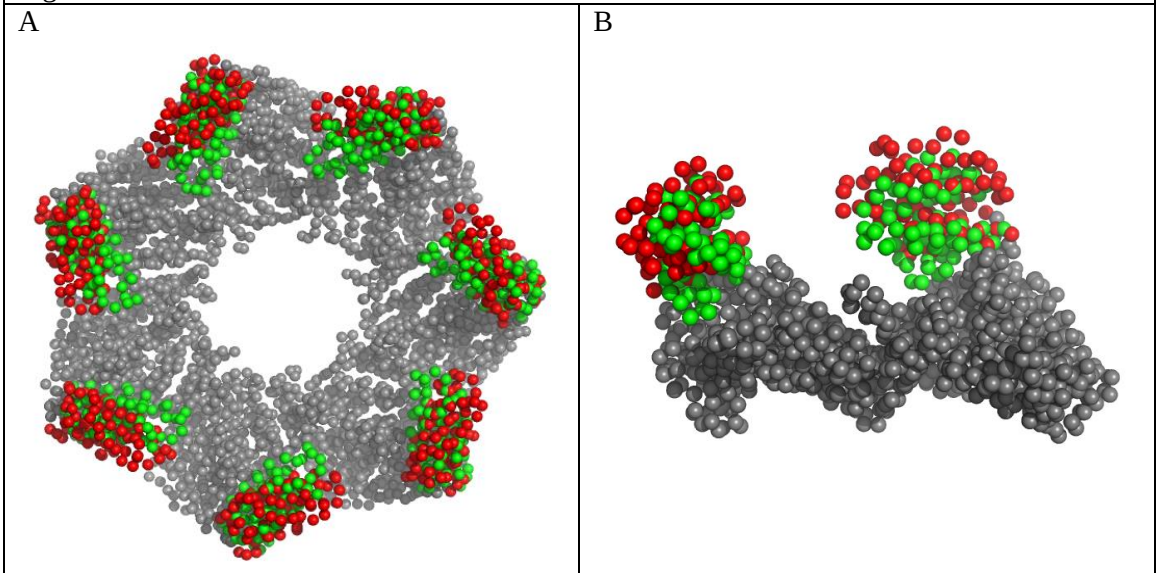
Intermediate Domain Movement	
Displacement	Rotation
A 	B 
Apical Domain Movement	
Displacement	Rotation
C 	D 
Displacement during the T → R transition of the intermediate (A) and apical (C) domains in the directions **radial**, **tangential** and **normal** subunit's internal coordinate system. Calculations were made using SVD as described in Sections 5.2.2.1 and 5.2.2.2.	Axis-angle rotation during the R → T relaxation of the intermediate (B) and apical (D) domains in the directions **radial**, **tangential** and **normal** subunit's internal coordinate system. Calculations were made using SVD as described in Sections 5.2.2.1 and 5.2.2.2.

Table 5.2 – Fitted Exponential Rate Terms for the Average Domain Movements That Characterize the T → R Transition

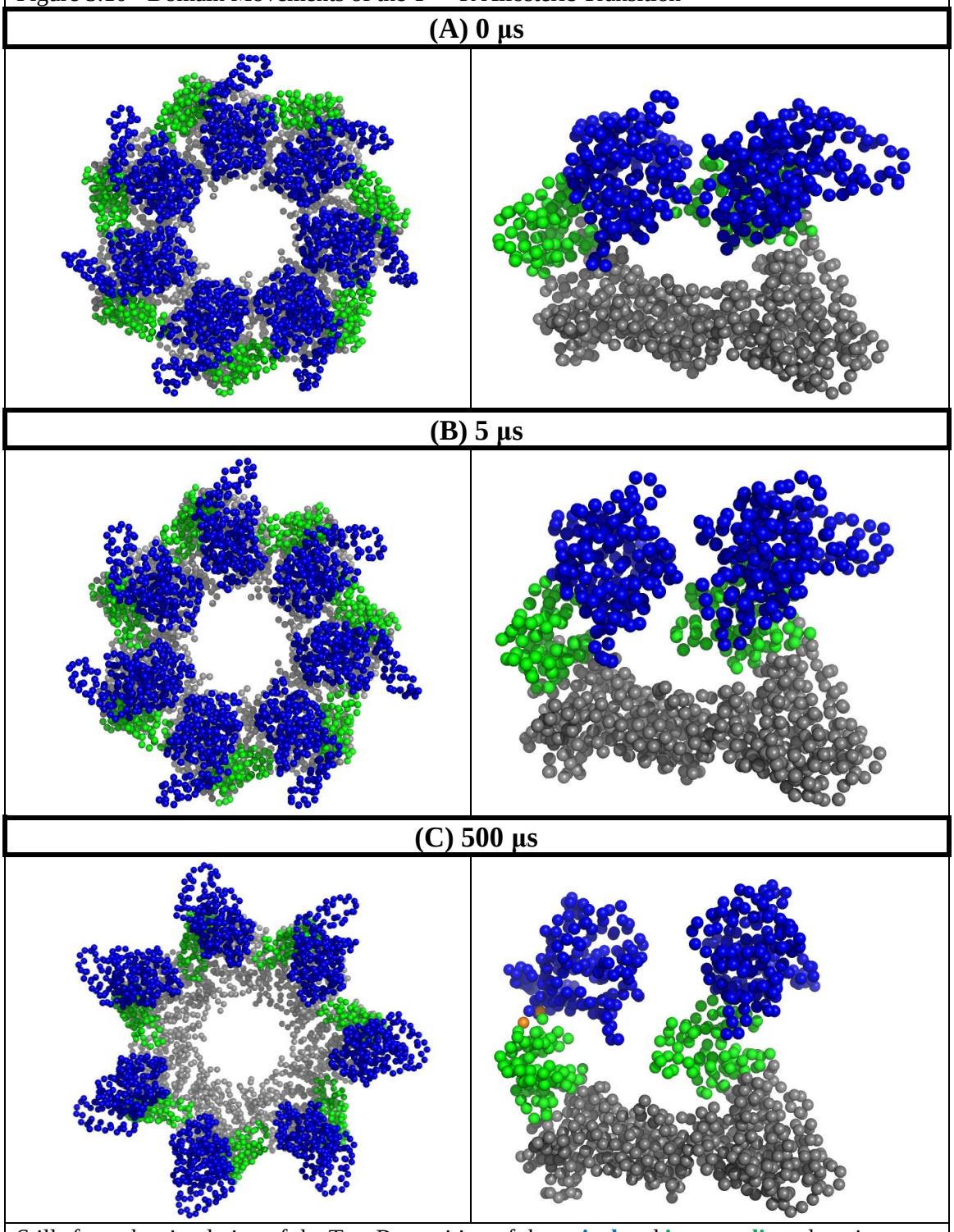
Intermediate Domain						
	Displacement			Rotation		
	X	Y	Z	X	Y	Z
k_1 (μs^{-1})	3.15 ± 0.06	3.08 ± 0.10	3.24 ± 0.08	3.51 ± 0.04	0.051 ± 0.018	3.24 ± 0.08
k_2 (μs^{-1})	33.1 ± 0.8	43 ± 3	240 ± 50	41.0 ± 0.6	28.1 ± 0.9	170 ± 22
k_3 (μs^{-1})		114 ± 7				
Apical Domain						
	Displacement			Rotation		
	X	Y	Z	X	Y	Z
k_1 (μs^{-1})	2.89 ± 0.02		2.82 ± 0.02	1.28 ± 0.02	0.02 ± 0.02	1.699 ± 0.019
k_2 (μs^{-1})	50. ± 5			5.37 ± 0.15	6.10 ± 0.09	12.1 ± 0.6

Figure 5.9 – Intermediate Domain Motion of the T → R Allosteric Transition



Superimposition of simulation frames of the intermediate domain during the T → R allosteric transition at **0 μs** and **500μs** after thermal equilibration. (A) Looking down the seven-fold axis of symmetry. The apical domain has been removed for clarity. (B) Looking normal to the seven-fold axis of symmetry.

Figure 5.10 - Domain Movements of the T → R Allosteric Transition



Stills from the simulation of the T → R transition of the **apical** and **intermediate** domains at the indicated **time points**. Views are looking down (left) and normal (right) to the seven-fold axis of symmetry. The highlighted orange bead in (C) corresponds to the hinge residue G374.

Close inspection of the final values for the extension of the apical domain radially and normal to the equatorial plane show that they are greater than the values calculated from the crystal structure of the ROpen end state structure by about 6Å (look ahead to the initial values of plots in Figure 5.11. at $t = 0$). That is, the domain movements observed during the simulated allosteric transition are larger than what would be expected based on the crystal structures. This additional movement is due to the lack of inter-domain and inter-subunit contacts in the apical domains of the ROpen structure. The apical domain is almost completely devoid of any native contacts with the intermediate domain, save a single flexible hinge residue, G374 (Figure 5.10.C in orange). Without any orienting native contacts, thermal motion causes the domains to then sample all available space. An equivalent expansion is observed during the control simulation of the R state.

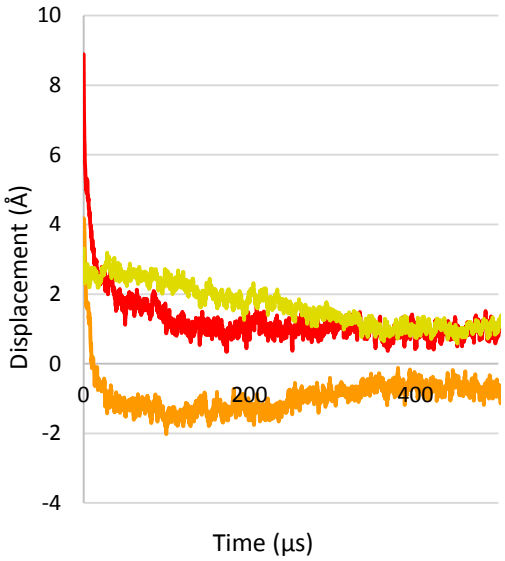
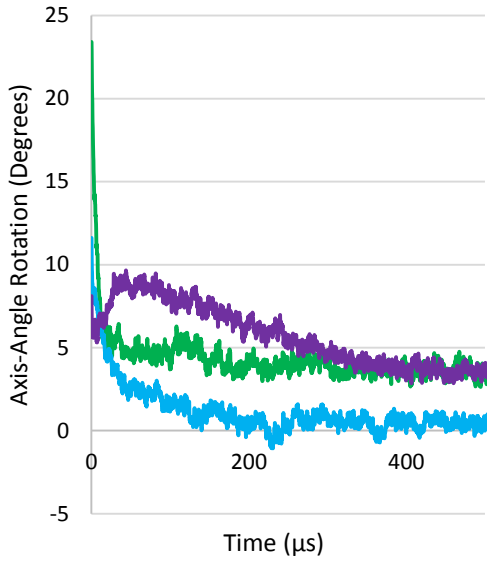
5.3.2.2 R → T Allosteric Relaxation

The ensemble average of each domain movement measure for the R → T state is given in Figure 5.11. Within the intermediate domain, the R →T transition appears to simply be the reverse of the T → R transition. The movements occur on comparable time scales and with very similar magnitudes. The rotation of the intermediate domain outward occurs in the fast phase. From this position, the domain is able to begin reestablishing contacts with the adjacent apical domain, which slowly relaxes from its extended position.

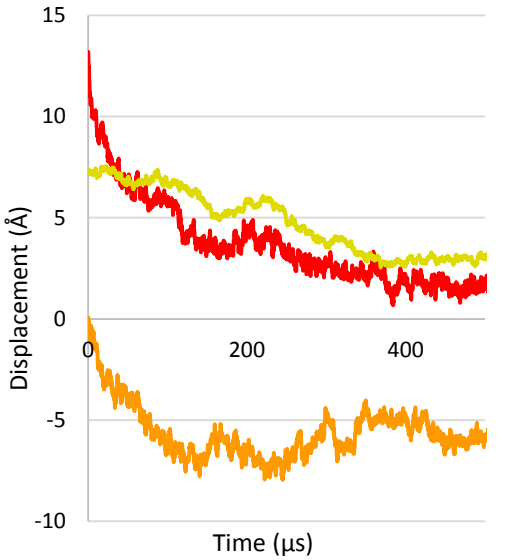
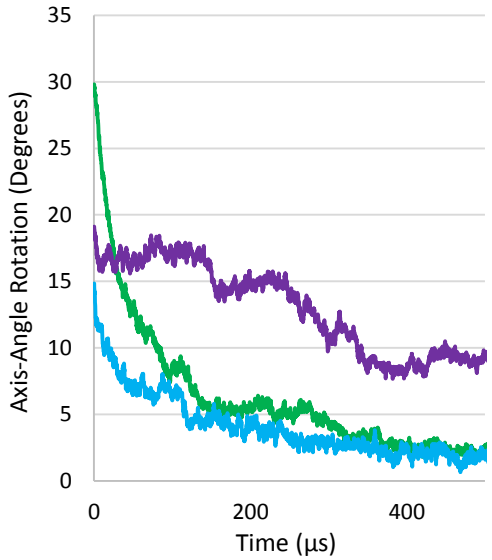
Unlike the intermediate domain measures, the average motions of the apical domain in the R → T transition are noisy curves that bear little resemblance to the smooth exponentials that described their movement during the T → R transition (Figure 5.8). Displacement and rotation around the the radial and tangential directions can be fit with double exponentials, revealing time scales similar to that of the T → R transition. In contrast, movement in the direction normal to the equatorial plane trends as expected, but the motion is poorly described as a sum of three or fewer

Figure 5.11 – Average Domain Movements of the R → T Allosteric Relaxation

Intermediate Domain Movement

Displacement	Rotation
A 	B 

Apical Domain Movement

Displacement	Rotation
C 	D 
Displacement during the R → T relaxation of the intermediate (A) and apical (C) domains in the directions **radial**, **tangential** and **normal** subunit's internal coordinate system. Calculations were made using SVD as described in Sections 5.2.2.1 and 5.2.2.2.	Axis-angle rotation during the R → T relaxation of the intermediate (B) and apical (D) domains in the directions **radial**, **tangential** and **normal** subunit's internal coordinate system. Calculations were made using SVD as described in Sections 5.2.2.1 and 5.2.2.2.

exponential terms. The plot instead appearing to decrease linearly until hitting some minimum value. In contrast, rotation around the normal axis appears to begin only after a delay period, then dropping off linearly until a floor value is reached. Thus, unlike their motion during the T→ R transition, the movement of individual apical domains within a ring is poorly characterized by the average of their movements. The analysis of ring symmetry in the next section provides an explanation for the observed behavior.

5.3.3 Ring Symmetry

In addition to comparing the movement of domains within a subunit, the symmetry of the domain movements within a ring is investigated. Ring asymmetry turns out to be a defining difference between the forward and reverse allosteric transitions simulated. If an allosteric transition is perfectly symmetric, then at any point in the simulation, the measures of any aspect of domain motion would be the same for all seven subunits within a ring. The deviation between measures of domain motion within a ring are indicative of transition asymmetry. Quantitating the degree of that deviation provides a measure of ring asymmetry. The analysis presented in this section is a comparison of the ring asymmetry observed during the T → R versus R → T allosteric transition. Ring asymmetry is plotted against a rough estimation of the transition progress, providing qualitative insight into the differences between the transitions.

5.3.3.1 Theory

The previous section demonstrated that SVD can be used to determine a set of six domain movement measures for each subunit within a ring, at each time-step in a given simulation trajectory. The total domain displacement within a subunit, say subunit A, can be calculated using Eq 5.8. The mean and standard deviation of the set of total displacement measures of a ring, $\mathcal{R} = \{r_i\}_{i=A}^{G}$, can be calculated using Eq 5.9 and 5.10, respectively. The standard deviation of the set is defined as the measure of asymmetry of the domains within a ring.

$$r_A = \sqrt{{r_{A,x}}^2 + {r_{A,y}}^2 + {r_{A,z}}^2} \qquad \text{Eq 5.8}$$

$$\langle \mathcal{R} \rangle = \frac{1}{7}\sum_{i=A}^{G} r_i \qquad \text{Eq 5.9}$$

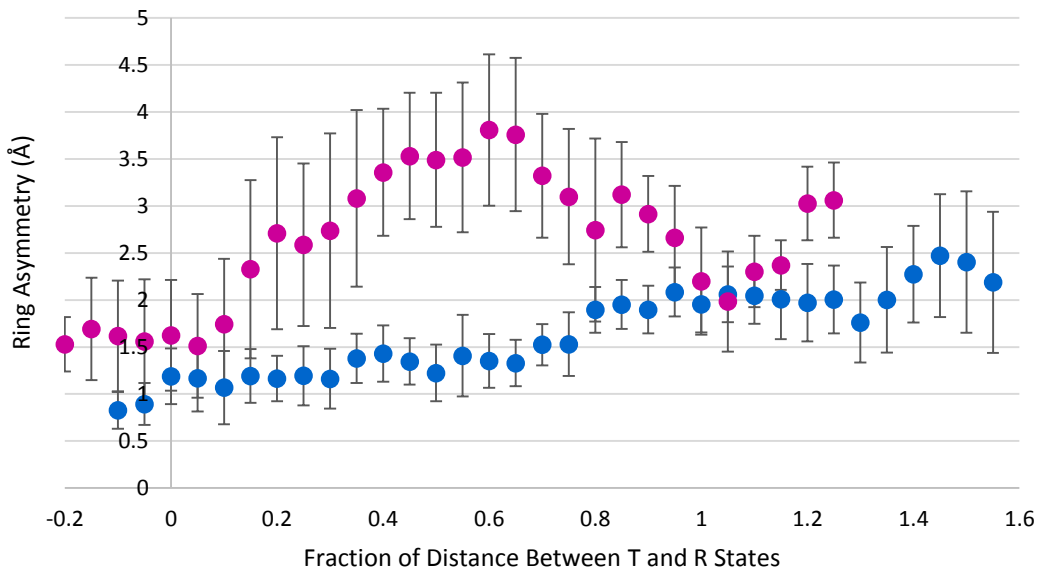
$$\sigma_{\mathcal{R}} = \sqrt{\frac{1}{7}\sum_{i=A}^{G} (r_i - \langle \mathcal{R} \rangle)^2} \qquad \text{Eq 5.10}$$

In order to facilitate comparison between the two different allosteric transitions simulated, the calculated mean domain displacement, $\langle \mathcal{R} \rangle$, was converted to a rough estimation of the percentage of transition progress. This allows us to compare ring symmetry at particular points along the T → R transition. Transition progress is estimated by assuming the domain movement proceeds linearly. That is, the allosteric transitions are assumed to proceed from their initial state to their final state along a straight line, and the position of the domain along that line is taken as a measure of transition progress. The mean domain displacement was converted to percentage transition progress using Eq 5.11, where $\bar{T}$ and $\bar{R}$ are the mean domain displacements of the standard simulations for the T and R state, respectively, described in Section 5.2.1.6.

$$P_{\mathcal{R}} = \frac{\langle \mathcal{R} \rangle - \bar{T}}{\bar{R} - \bar{T}} \qquad \text{Eq 5.11}$$

Each output of a given simulation trajectory is converted to the pair of coordinates $(P_{\mathcal{R}}, \sigma_{\mathcal{R}})$ using the method described above. The procedure is repeated for all outputs within a given trajectory and across 15 simulation trials. The results are compiled into a single set of coordinate pairs, which is further reduced by binning $P_{\mathcal{R}}$ values. The ring asymmetry measures, $\sigma_{\mathcal{R}}$, that correspond to binned $P_{\mathcal{R}}$ values are averaged and the standard deviation is calculated as a measure of error.

Figure 5.12 – Asymmetry in the Allosteric Transitions of GroEL



Plots of apical domain asymmetry within a ring as a function of the progress made between the T and R states. Asymmetry and transition progress are calculated according to Section 5.3.3.1. Transition progress of 0 is considered the T state and 1 is considered the R state. Position of apical domains within a ring during the **T → R allosteric transition** is relatively symmetric, indicating a concerted transition. In contrast, the simulated **R → T allosteric relaxation** (going leftward, from 1 to 0), is highly asymmetric for most all of the transition.

5.3.3.2 Results

Analysis of intra-ring symmetry of the apical domain during the T → R and R → T allosteric transitions are presented in Figure 5.12. The T → R transition appears to occur in a relatively concerted fashion, consistent with previous findings. This is indicated by the slow initial build of ring asymmetry and the relatively low error of each data point along the T → R transition. After completing ~75% of the allosteric transition, ring asymmetry begins to build more rapidly and error of each point begins to increase. As was explained above, the T → R transition proceeds beyond the R state, as defined by the ROpen structure, because its only remaining inter-domain native contact is with the hinge residue G374, allowing it to freely explore additional configurations.

The R → T allosteric relaxation in Figure 5.12 stands in stark contrast to forward transition. Since it is the reverse transition, the system initially begins in the R state and allosteric transition

progress precedes leftward on the graph. Ring asymmetry immediately beings to increase as the transition proceeds, nearly doubling before finally reaching a peak. Furthermore, the error associated with each data point is relatively high, about double the size of the corresponding T → R points. After the transition is ~50% complete, asymmetry begins to slowly decrease, only reaching a value consistent with the asymmetry of the R state at ~15%. The plots indicate that the R → T transitions simulated proceed through a highly asymmetric set of intermediate states. Nearly 80% of a subunit's journey (or at least a subset of subunits within each ring) from the R back to the T state appears to occur independent of the other subunits within a ring.

In theory, the same analysis could be applied to the intermediate domain as well, but the speed with which the domain transitions proceed make the resulting data unreliable. For both allosteric transitions, 80% of the domain movement occurs over less than 5% of $(P_{\mathcal{R}}, \sigma_{\mathcal{R}})$ coordinate pairs. The speed of the transition also suggest that the results of such an analysis may be misleading because any observed symmetry may not be driven by inter-subunit contacts, but by the symmetric application of changes to the system's Hamiltonian.

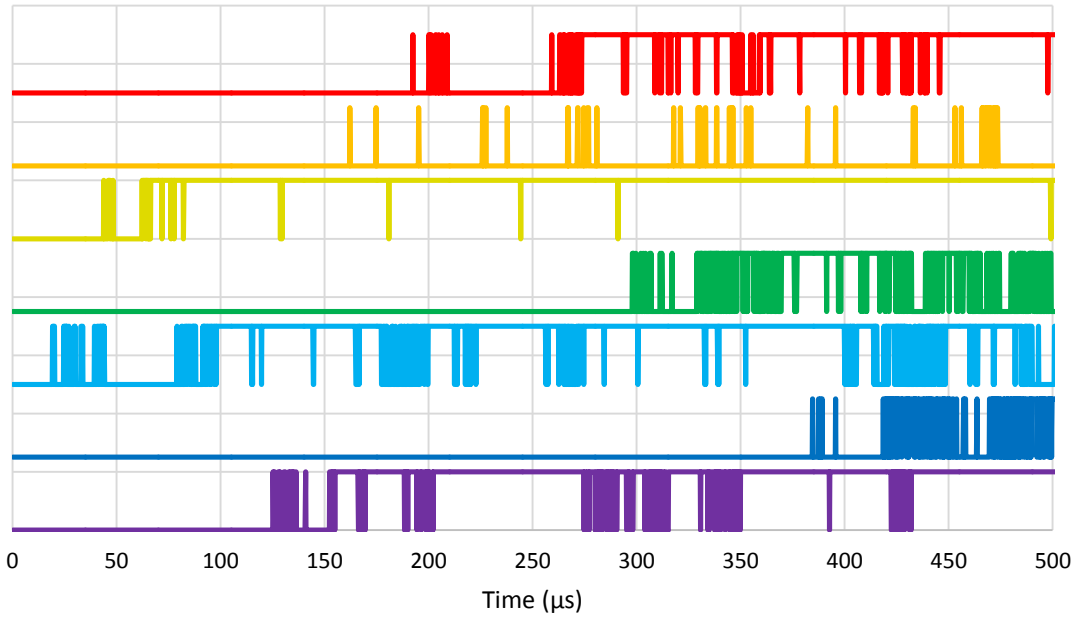
5.3.4 Subunit Docking

The proceeding analysis has led to the surprising result that, while the T → R transition displays a high degree of concerted motion, the R → T allosteric relaxation is a highly asymmetric process, with ring symmetry only being reestablished after the vast majority of the transition has been completed. The results seem to imply the subunits are able to adopt some T-like confirmation that is relatively asymmetric. Since the concerted motion model is ruled out from the analysis of the previous section, this section explores the R → T allosteric relaxation within the context of a stochastic model of subunit relaxation.

5.3.4.1 Theory

In order to further investigate the R → T allosteric relaxation simulated we will utilize the salt

Figure 5.13 – Inter-Subunit Docking of D216/K226 During the R → T Allosteric Relaxation

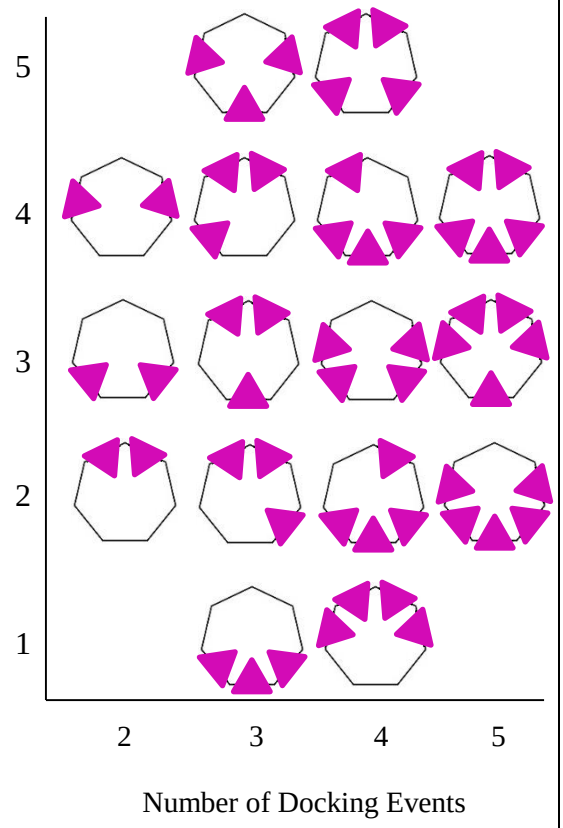


This is a plot of the docking of subunits measured by the D216/K226 salt bridge formation. If the seven subunits of GroEL are arbitrarily labeled A-G, then the plots correspond to the docking of **AB**, **BC**, **CD**, **DE**, **EF**, **FG** and **GA**. At t = 0, all subunits are undocked in the R state. Docking between subunit pairs is indicated by a unit jump in the relevant curve. Curves are offset from one another for clarity.

bridge measurements described in Section 5.2.2.3 and listed in Table 5.2. As the subunits within a ring begin to relax, inter-subunit salt bridge contacts begin to reform. Reestablishing these contacts presumably help to drive the allosteric relaxation down to the final, compact, T state. Formation of inter-subunit salt bridges is examined by monitoring the distance between the various charged pairs. The residues are flagged if they come into close enough contact to form a salt bridge. Since all charged pairs considered are either Glu/Lys or Glu/Arg, assuming that 4Å is the maximal distance between charge groups that constitutes an electrostatic bond, the maximal C_α–C_α distance between them that indicates interaction is 14.2Å, calculated using a simple ball-and-stick geometric model of each residue. The term "subunit docking" is used to describe the formation of an inter-subunit salt bridges throughout the chapter.

A quick look at the subunit docking results for a few individual trajectories suggests that the allosteric transition being studies is not a simple relaxation process. An example trajectory for the D216/K226 salt bridge is given in Figure 5.13. In the example given, a few subunits (light blue,

Figure 5.14 – Illustration of Docking States



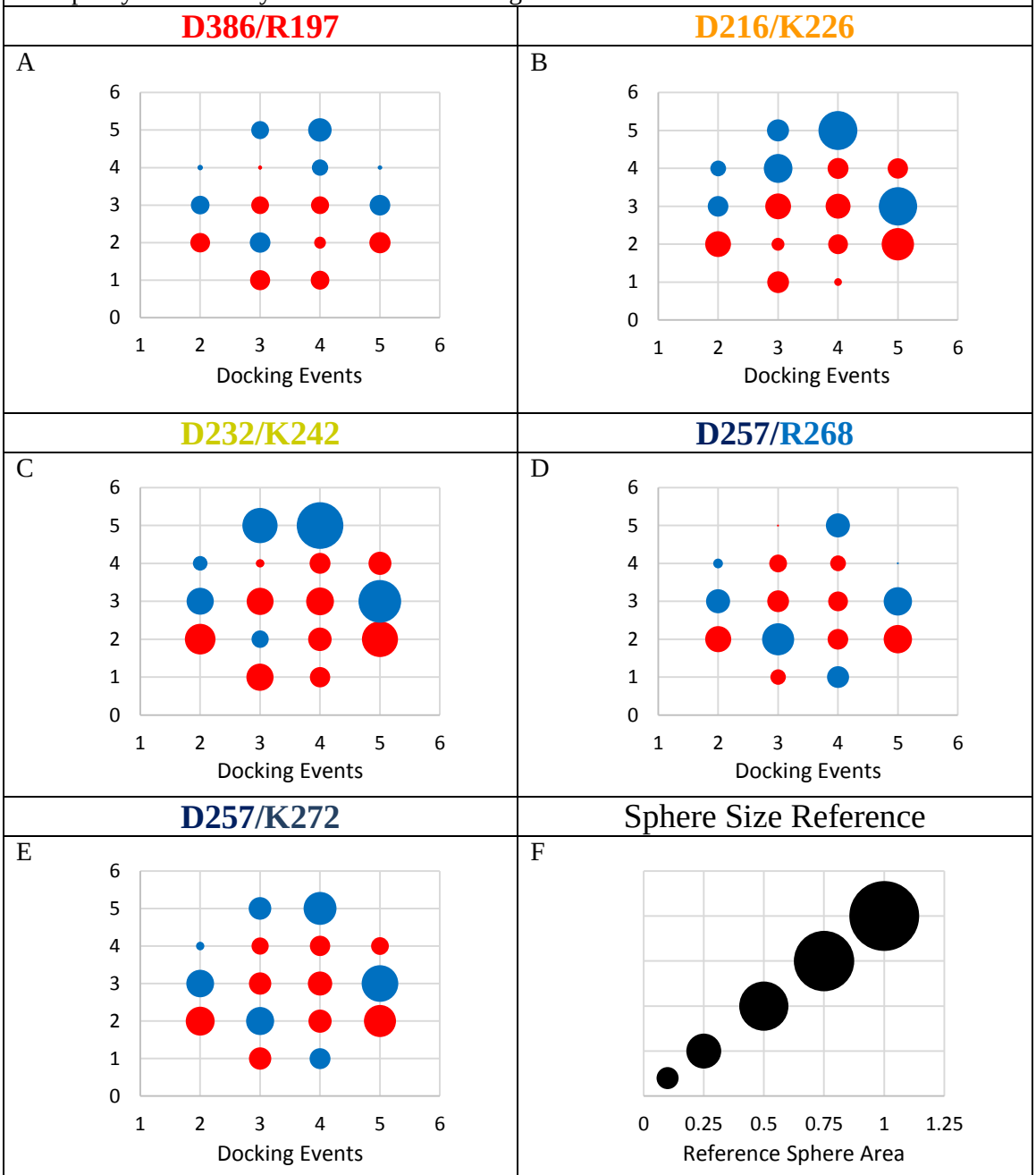
The set of all possible distinct docking states along the path from the R state (0 docking events) to the T state (7 docking events) during the R → T allosteric relaxation. **GroEL** is represented as a hexagon with subunits positioned at the seven corners. **Docking events** occur between subunits. Each column contains the set distinct configurations, or "docking states", with the indicated number of **docking events**. If docking is stochastic, all configurations in a column are equally likely. There is only one possible configuration for the docking states corresponding to 0,1,6 and, 7docking events, so they are omitted. The numeric labeling on the y-axis is arbitrary and used again in Figure 5.15 for easy docking configuration reference.

purple and especially yellow) appear to be stably docked to one another while others appear to bind less stably (red and orange), performing numerous reversible docking events. The remaining subunits (green and especially dark blue) appear as though an opportunity to dock wasn't available until nearly the end of the simulation. The results suggests that the observed ring asymmetry is caused by a dynamic subunit docking event that occurs during the final stages of the allosteric relaxation.

The results in Figure 5.13 illustrate another curious aspect of the subunit docking process: the strongly docked subunits (light blue, purple and yellow) are each spaced at least one subunit away from one another. This seems to indicate stability of a particular docking configuration, demonstrating that the relaxation of subunits is likely not stochastic. Instead, it appears that the binding of a first subunit makes a second docking event less favorable.

In the stochastic model of subunit relaxation, as the subunits within a ring return to the T state, they form their inter-subunit salt bridges in a random and independent manner. The stochastic model allows us to predict the distribution of docked states, which is going to look identical to the distribution of dye labeled states in Chapter 1 and re-depicted in Figure 5.14. Only one distinct

Figure 5.15 – Sphere Plot of Differences Between Experimental Docking State Occupancy and Occupancy Predicted by the Stochastic Docking Model of the R → T Allosteric Relaxation



Comparison of docking state occupancy, as determined by the indicated salt bridge pair (coloring consistent with Table 5.1). Each sphere corresponds to a distinct docking state (arbitrarily numbered and consistent with Figure 5.14) within a column of total docking events. The area of each sphere is proportional to the difference between measured and predicted occupancy, and colored depending on if that difference is **positive** or **negative**. In other words, the color coding indicates whether a state is **more probable** or **less probable** than the stochastic prediction. Panel (F) contains a set of reference spheres.

configuration exists for the states corresponding to 0,1,6 and, 7 docking events, so they are omitted from the analysis. The population of each docked state that corresponds to a particular number of total docking events, i.e. the states within a column in Figure 5.14, should occur with

equal probability if the process is stochastic. Thus measuring the probability of occupancy of a particular docking state provides us with a simple way to check the stochastic docking assumption.

For a given simulation trajectory, each salt-bridge pair within a ring was analyzed and scored as either docked or not docked. This was performed for all outputs of all simulation trajectories of the R → T allosteric relaxation. These results were then put through a program that processed them into docking configurations, assigning them to one of the 20 distinct structures in Figure 5.14. Once finalized, the occupancy probability was calculated and compared to the predicted value assuming stochastic docking.

5.3.4.2 Results

The difference between the observed and predicted occupancy probabilities are displayed in Figure 5.15 for the 5 inter-subunit salt bridges monitored. The results of Figure 5.15 demonstrate that the docking of subunit salt-bridges that concluded the R → T allosteric relaxation is far from a stochastic process. Red dots indicate states that are less favorable than the stochastic model, whereas blue indicates states that are more favorable. The blue states then trace out the most probable path through docking events for the allosteric transition. In general, the favorable states for all salt bridge pairs are the ones where docked subunits are pairwise evenly distributed across the ring, indicating that multiple binding events in a row, on the same subunit, are generally unfavorable. This suggests the docking of subunits within a ring is negatively cooperative, such that the docking of one subunit will make the docking of the next subunit less favorable.

5.4 CONCLUSION

The motion of GroEL's subunits during its allosteric transitions are often described by the rigid motion of two domains: the intermediate and apical domains. However, this characterization has

yet to be quantitatively demonstrated throughout the course of the allosteric transition. SVD analysis was applied to the MD simulation trajectories in order to retrieve the best-fit rigid transform. The results were then compared to the *in silico* obtained experimental data to obtain a quantitative analysis of the rigid motion assumption. This analysis showed that subunits in both the forward and reverse allosteric transitions are well-described by the rigid body movements of its two domains. This allows us to reduce the observed motion in each simulation output, which contains over 14,000 3D coordinates, into two sets of six domain measures.

The well-established concerted nature of the T → R allosteric transition has influenced years of GroEL research. The idea of concerted motion has been applied, by assumption, to numerous other steps in the catalytic cycle. The application of this assumption to several catalytic steps was explored in the previous three chapters. Specifically, Chapter 2 demonstrated that the release of ADP from the asymmetric complex is a multi-step process, implicating the importance of the R → T allosteric relaxation in governing the rate of cycling in the system. The current chapter utilized MD simulations of the forward and reverse allosteric transitions to explore the differences between the two allosteric transitions. These simulations demonstrated that the allosteric relaxation is not simply the reverse of the T → R allosteric transition, but instead a highly asymmetric relaxation.

After reducing the number of variables used to describe a given allosteric transition, the average value of each domain movement measure can be calculated. For the T → R transition, the motion of each domain was shown to be well-approximated by the averaged motion of all subunits across multiple simulations. The average domain measures can be analyzed to reveal the magnitudes and time scales of the motion. The transition path observed was in good agreement with previous studies. A slight rotation of the intermediate domain is followed by a pronounced extension and rotation of the apical domain, expanding the central cavity and increasing the distance between SP binding sites.

Intermediate domain motion in the R → T allosteric relaxation was found to be well-approximated as the simple reverse of the T → R transition. In contrast, the apical domain motion of the R → T allosteric relaxation was not well characterized by the average of its motion. This was observed despite averaging over 15 different simulations, compared with 5 for the T → R transition. The results suggest that, intra-ring asymmetry during the allosteric relaxation is disordered and/or follows a different trajectory than the forward transition.

Further analysis of the domain movements demonstrate that the T → R transition is concerted within a ring, a result supported by a large body of previous research. Not only is asymmetry within a ring low (indicated by the low value of the asymmetry measure in Figure 3.12), each T → R transition simulated occurs along the same path, indicated by the low variation, or error, of asymmetry measures at a given point throughout the transition. As expected, once the R state is reached, asymmetry within a ring begins to increase due to the lack of inter-subunit contacts.

Unlike the forward allosteric movements, the R → T allosteric relaxation is highly asymmetric throughout the majority of the transition. Asymmetry actually increases as the subunits begin to relax and variations between simulations are quite large. Domain symmetry within a ring enters the "concerted territory" of the T → R transition after ~85% of the relaxation is complete. The analysis undermines the assumption that the R → T transition is comprised of concerted domain movements.

The final portion of this chapter explored the mechanism of the R → T allosteric relaxation. The preceding section had demonstrated that the relaxation was not concerted. Another possibility, suggested by the lack of inter-subunit contacts, is a stochastic mechanism in which each subunit relaxes back to the T state independently of the others within a ring. This hypothesis was tested by monitoring the distribution of species formed during the course of the allosteric relaxation. The analysis presented in Section 5.3.4 demonstrates that the formation of salt bridges that dock

the apical domain of subunits together is not a stochastic process, but favorably occurs within only a small subset of possible states. The states preferably occupied all display an even spread of docked subunits within a ring, indicating that the binding of two subunits makes the docking of a third less favorable. Thus, the allosteric relaxation simulated displays dynamic negative-cooperativity as the ring makes its way back down to the T state.

Although contrary to popular belief, the idea of a slow dynamically cooperative allosteric relaxation is consistent with much of what is already well established in the GroEL/GroES catalytic cycle. In the absence of SP, ADP release is the rate limiting step in the cycle, perhaps because it's closely coupled to relaxation. Furthermore, the results from Chapter 2 of this work demonstrate that the ADP release portion of the cycle is a multi-step process. Previous single molecule experiments identified a large population of T-like structures that exist under conditions of saturating ATP, a phenomenon that could feasibly be attributed to a population of relaxation intermediates. The results are even consistent with GroEL's responsiveness to temperature, as an increase in thermal motion would prevent the formation of the semi-stable pair-wise docked states that slow the rate of relaxation.

Perhaps most importantly, a dynamically cooperative allosteric relaxation has mechanistic implications as to how SP is able to increase the rate of nucleotide exchange. The binding of SP to multiple apical domains would effectively place the ring into one of the unfavorable states with a run of contiguously docked subunits, while simultaneously blocking access to the pair-wise docked states. This higher energy configuration would then have a clear, down-hill ride to the fully docked state.

The degree to which the MD simulations performed in this chapter accurately reflect the true allosteric relaxation in the GroEL/GroES system has yet to be shown. Nonetheless, hopefully this and previous chapters in this dissertation have made it clear that the widely held dogma of the

GroEL ring acting as a unified concerted machine is in need of revision, and that its asymmetric behavior may play just as important of a role in governing the various facets of its catalytic cycling.

5.5 REFERENCES

1. Todd, M. J., G. H. Lorimer and D. Thirumalai. "Chaperonin-Facilitated Protein Folding: Optimization of Rate and Yield by an Iterative Annealing Mechanism." *Proc Natl Acad Sci U S A* 93, no. 9 (1996): 4030-5.

2. Lin, J. C. and D. Thirumalai. "Relative Stability of Helices Determines the Folding Landscape of Adenine Riboswitch Aptamers." *J Am Chem Soc* 130, no. 43 (2008): 14080-1.

3. Corsepius, N. C. and G. H. Lorimer. "Measuring How Much Work the Chaperone Groel Can Do." *Proc Natl Acad Sci U S A* 110, no. 27 (2013): E2451-9.

4. Motojima, Charu Chaudhry, W. A. Fenton, G. W. Farr and A. L. Horwich. "Substrate Polypeptide Presents a Load on the Apical Domains of the Chaperonin Groel." *PNAS* 101, no. 42 (2002): 15005.

5. Grason, J. P., J. S. Gresham, L. Widjaja, S. C. Wehri and G. H. Lorimer. "Setting the Chaperonin Timer: The Effects of K+ and Substrate Protein on Atp Hydrolysis." *Proceedings of the National Academy of Sciences* 105, no. 45 (2008): 17334-17338.

6. Grason, J. P., J. S. Gresham and G. H. Lorimer. "Setting the Chaperonin Timer: A Two-Stroke, Two-Speed, Protein Machine." *Proc Natl Acad Sci U S A* 105, no. 45 (2008): 17339-44.

7. Sameshima, T., R. Iizuka, T. Ueno and T. Funatsu. "Denatured Proteins Facilitate the Formation of the Football-Shaped Groel-(Groes)2 Complex." *Biochem J* 427, no. 2 (2010): 247-54.

8. Yang, D., X. Ye and G. H. Lorimer. "Symmetric Groel:Groes2 Complexes Are the Protein-Folding Functional Form of the Chaperonin Nanomachine." *Proc Natl Acad Sci U S A* 110, no. 46 (2013): E4298-305.

9. Ye, X. and G. H. Lorimer. "Substrate Protein Switches Groe Chaperonins from Asymmetric to Symmetric Cycling by Catalyzing Nucleotide Exchange." *Proc Natl Acad Sci U S A* 110, no. 46 (2013): E4289-97.

10. Ma, J., P. B. Sigler, Z. Xu and M. Karplus. "A Dynamic Model for the Allosteric Mechanism of Groel." *J Mol Biol* 302, no. 2 (2000): 303-13.

11. Kipnis, Y., N. Papo, G. Haran and A. Horovitz. "Concerted Atp-Induced Allosteric Transitions in Groel Facilitate Release of Protein Substrate Domains in an All-or-None Manner." *Proceedings of the National Academy of Sciences* 104, no. 9 (2007): 3119-

3124.

12. Papo, N., Y. Kipnis, G. Haran and A. Horovitz. "Concerted Release of Substrate Domains from Groel by Atp Is Demonstrated with Fret." *J Mol Biol* 380, no. 4 (2008): 717-25.

13. Chaudhry, Charu, Arthur L. Horwich, Axel T. Brunger and Paul D. Adams. "Exploring the Structural Dynamics of the E.Coli Chaperonin Groel Using Translation-Libration-Screw Crystallographic Refinement of Intermediate States." *Journal of Molecular Biology* 342, no. 1 (2004): 229-245.

14. Danziger, O., D. Rivenzon-Segal, S. G. Wolf and A. Horovitz. "Conversion of the Allosteric Transition of Groel from Concerted to Sequential by the Single Mutation Asp-155 -> Ala." *Proc Natl Acad Sci U S A* 100, no. 24 (2003): 13797-802.

15. Hyeon, G. H. Lorimer and D. Thirumalai. "Dynamics of Allosteric Transitions in Groel." *PNAS* 103, no. 50 (2006): 18939.

16. Sorkine, O. "As-Rigid-as-Possible Surface Modeling." *AMC International Conference Proceeding Series* 257, (2007): 109-116.

17. Boisvert, D. C., J. Wang, Z. Otwinowski, A. L. Horwich and P. B. Sigler. "The 2.4 a Crystal Structure of the Bacterial Chaperonin Groel Complexed with Atp Gamma S." *Nat Struct Biol* 3, no. 2 (1996): 170-7.

18. Ranson, N. A., G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich and H. R. Saibil. "Atp-Bound States of Groel Captured by Cryo-Electron Microscopy." *Cell* 107, no. 7 (2001): 869-79.

19. Saibil, H. R., W. A. Fenton, D. K. Clare and A. L. Horwich. "Structure and Allostery of the Chaperonin Groel." *J Mol Biol* 425, no. 9 (2013): 1476-87.

20. Fei, X., D. Yang, N. LaRonde-LeBlanc and G. H. Lorimer. "Crystal Structure of a Groel-Adp Complex in the Relaxed Allosteric State at 2.7 a Resolution." *Proc Natl Acad Sci U S A* 110, no. 32 (2013): E2958-66.

21. Mickler, M., R. I. Dima, H. Dietz, C. Hyeon, D. Thirumalai and M. Rief. "Revealing the Bifurcation in the Unfolding Pathways of Gfp by Using Single-Molecule Experiments and Simulations." *Proc Natl Acad Sci U S A* 104, no. 51 (2007): 20268-73.

22. Dima, R. I. and H. Joshi. "Probing the Origin of Tubulin Rigidity with Molecular Simulations." *Proc Natl Acad Sci U S A* 105, no. 41 (2008): 15743-8.

23. Hyeon, C. and D. Thirumalai. "Mechanical Unfolding of Rna: From Hairpins to Structures with Internal Multiloops." *Biophys J* 92, no. 3 (2007): 731-43.

24. Lin, J. C. and D. Thirumalai. "Relative Stability of Helices Determines the Folding Landscape of Adenine Riboswitch Aptamers." *J Am Chem Soc* 130, no. 43 (2008): 14080-1.

25. Hyeon, C., P. A. Jennings, J. A. Adams and J. N. Onuchic. "Ligand-Induced Global

Transitions in the Catalytic Domain of Protein Kinase A." *Proc Natl Acad Sci U S A* 106, no. 9 (2009): 3023-8.

26. Schlick, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Vol. 21. 2 ed. New York: Springer, 2010.

27. Betancourt, M. R. and D. Thirumalai. "Exploring the Kinetic Requirements for Enhancement of Protein Folding Rates in the Groel Cavity." *J Mol Biol* 287, no. 3 (1999): 627-44.

28. Braig, K., Z. Otwinowski, R. Hegde, D. C. Boisvert, A. Joachimiak, A. L. Horwich and P. B. Sigler. "The Crystal Structure of the Bacterial Chaperonin Groel at 2.8 A." *Nature* 371, no. 6498 (1994): 578-86.

29. Xu, Z., A. L. Horwich and P. B. Sigler. "The Crystal Structure of the Asymmetric Groel-Groes-(Adp)7 Chaperonin Complex." *Nature* 388, no. 6644 (1997): 741-50.

30. Ranson, N. A., G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich and H. R. Saibil. "Atp-Bound States of Groel Captured by Cryo-Electron Microscopy." *Cell* 107, no. 7 (2001): 869-79.

31. Murai, N., Y. Makino and M. Yoshida. "Groel Locked in a Closed Conformation by an Interdomain Cross-Link Can Bind Atp and Polypeptide but Cannot Process Further Reaction Steps." *J Biol Chem* 271, no. 45 (1996): 28229-34.

32. Tehver, R., J. Chen and D. Thirumalai. "Allostery Wiring Diagrams in the Transitions That Drive the Groel Reaction Cycle." *J Mol Biol* 387, no. 2 (2009): 390-406.

33. Lay, D. *Linear Algebra and Its Applications*. Vol. 3. Boston: Pearson, 2006.

Chapter 6

Dissertation Summary

6.1 CONCLUSION AND PROPOSED FUTURE DIRECTION

6.1.1 Quantitating the Impact of Substrate Protein

TMR dimers were shown to be suitable spectroscopic probes to report the allosteric state of GroEL The use of dye dimers to report conformational changes in proteins has been previously demonstrated (1-3), and can potentially be applied to a variety of enzymatic systems as a sensitive spectroscopic method for monitoring distance changes that are too subtle for FRET based probes to detect. Absorbance spectroscopy can also be used to monitor dye dimer populations, offering researchers an alternative measure that is more robust and can be utilized in experimental conditions that do not permit quantitative fluorescence measurements.

Additionally, TMR dimers were shown to mimic the effect of SP on the GroEL catalytic cycle, increasing the mean cycle time (4-8). The finding supports the view that the non-covalent tethering of subunits within a ring leads to the observed acceleration in ADP release (9, 10). Utilizing TMR dimers as an SP surrogate circumvents the complications normally associated with studying the role of SP in the chaperonin catalytic cycle. Modification of existing computational framework describing the activity of GroEL led to the first quantitative description of the role SP plays in shifting the allosteric equilibrium.

Future research focused on the development of various dye dimer systems as spectroscopic probes would serve to expand the tool chest of chemistry and biochemistry researchers. By varying the fluorophore and/or linker molecules, there exists the potential to generate a library of self-reporting molecular latches that vary in degrees of associative stability. Development of such a collection could afford researchers with a technique for measuring work that utilizes inexpensive and readily available spectroscopic techniques, rather than the currently available methods that require expensive atomic force microscopes.

The results of Chapter 1 also have a significant impact within the field of GroEL/GroES research. Whether GroEL is an active or passive player in catalyzing the folding of its SPs is one of the most hotly contested topics among GroEL/GroES researchers (11-13). The experiments of Chapter 1 conclusively demonstrate the ability of the T → R allosteric transition to perform work through its domain movements. The results lend support to the iterative annealing model of GroEL function, which hypothesize the importance of GroEL's ability to forcible unfold bound SP in order to catalyze productive folding.

6.1.2 Asymmetric Allosteric Relaxation

The pre-steady state kinetic analysis of Pi release in Chapter 2 identified an unknown species that is significantly populated throughout the course of the GroEL/GroES catalytic cycle. Subsequent analysis demonstrated the ability to conclusively place the species in the ADP release portion of the cycle. These results were corroborated by the MD simulations presented in Chapter 5, which suggest the importance of the R → T allosteric relaxation in governing the catalytic dynamics. The simulations showed that subunits relaxed to the T state by forming pairwise inter-subunit interactions that made the docking of a third subunit less favorable. In other words, allosteric relaxation of subunits displayed dynamic negative cooperativity. This supports the pre-steady state kinetic results of Chapter 2, suggesting that the multi-step ADP release process observed is a result of the formation of quasi-stable asymmetric intermediate configurations within a ring (14).

The results of Chapters 2 and 5 are significant because they deviate from the existing theory that the domain movements of the allosteric transitions in the GroEL are concerted within a ring (15, 16). The findings support a recent observation of an asymmetric ring species (17) and hypothesize the importance of asymmetry in governing catalytic cycling. These results also suggest the importance of the asymmetric allosteric relaxation in allowing for the system to dynamically respond to the needs of the cell. In the absence of SP, the asymmetric relaxation

governs the slow release of ADP. When folding catalysis is in high demand, SP binding prevents the pairwise docking of subunits, stimulating efficient relaxation that leads to more rapid turnover.

Additional studies are needed to conclusively demonstrate the proposed model of dynamic negative cooperativity in the allosteric relaxation. The model would be supported by the observation of the pairwise docked intermediates. One potential experimental method could be to perform a series of inter-subunit cross-linking experiments, using UV triggered cross-linking agents. Performing the experiments in the presence and absence of SP could potentially produce the distributions of docking configurations that would illustrate the validity of the model.

6.1.3 Stochastic Mechanism of ATP Hydrolysis

Chapters 2 and 3 offer a mathematically rigorous framework for analyzing the pre-steady state kinetics of an enzymatic system. The method presented allowed for the extraction of a large amount of information from a limited number of experimental data sets. This is due to the model specific nature of the analysis and the care taken to disprove alternative models. The structure of the presented study could potentially be applied to any number of enzymatic systems, allowing for an accurate formulation of the enzyme's catalytic model.

A noteworthy result obtained from the analysis in Chapter 2 was that ATP hydrolysis could be defined as a single step process. The error breakdown presented in the chapter demonstrated how, out of all steps in the proposed catalytic model, we can be the most confident about accurately describing the hydrolysis kinetics. This is in contrast to the analysis in Chapter 3, which demonstrated that hydrolysis must occur through a series of species. The two observations of the same system are only reconciled if the mechanism of ATP hydrolysis in the GroEL/GroES system is stochastic. The result is important because it departs from the previously proposed theory that ATP hydrolysis is concerted within a GroEL ring (18, 19). This allows for a more

accurate description of the processes that govern the catalytic cycling of GroEL/GroES and protein folding catalysis.

The results of the pre-steady state kinetic analysis presented in Chapters 2 and 3 rely on a single technique for measuring free Pi and the population of $GroEL_7$/GroES complex, respectively. Any additional experiments that utilize other methods for measuring these observables would allow for greater confidence in the results. Similarly, the findings would be strengthened by any additional experiments that monitored one of the other system observables: free ATP, free ADP, the ATP bound species, the symmetric species, the population of species in each allosteric state, etc. In general, multiple observations of the same enzymatic system are necessary to build a complete and accurate catalytic model.

6.1.4 Breakage of GroEL/$GroES_2$ Complex Symmetry

Chapters 2 and 3 established that ATP hydrolysis is stochastic and that it occurs through series of symmetric GroEL/$GroES_2$ species. This led us to question the existing theory of GroES exchange in Chapter 4, which states that all seven ATP within a ring must be hydrolyzed before GroES dissociation. It was first demonstrated that GroES dissociation is a time-independent process. This is significant because it implies there is some conformational switch that occurs in the GroEL/$GroES_2$ complex structure that allows for GroES release. The subsequent analysis presented in the chapter demonstrates that the GroES dissociation from a ring is not only dependent on the number of ATP's hydrolyzed within the ring, but also the number of hydrolysis events that have occurred on the adjacent ring.

The breakage of symmetry model proposed in Chapter 4 identifies which nucleotide states form stable symmetric complexes, and which promote the dissociation of GroES. The analysis relies on the stochastic hydrolysis mechanism and the observed decay of the population of symmetric species as a function of observed hydrolysis events. Additional support for the model could be

obtained by using high resolution native mass spectrometry. This would allow for direct observation of the distribution of nucleotide states in the symmetric and asymmetric particles. Use of the subunit mixed samples of heavy D398A and light wild-type GroEL would help to exaggerate the mass difference and improve resolution of the various nucleotide states.

Although the analysis in Chapter 4 was able to demonstrate the existence of a structural switch and identify the nucleotide configurations that lead to the dissociation of GroES from the symmetric GroEL/$GroES_2$ complex, there is no evidence of the mechanistic nature of the configuration change that triggers GroES release. Future experiments are needed to identify the network of residues responsible for coupling the inter- and intra- ring signals that lead to the breakage of symmetry. Point mutation studies on GroEL could be used to identify the responsible residue(s). MD simulations may also play a role in their identification.

6.1.5 Concluding Remarks

The analysis presented in this dissertation addresses a number of aspects in the GroEL/GroES catalytic cycle that are often overlooked or oversimplified in the existing literature. The approach employed throughout the five chapters is to analyze sets of experimental data using a rigorously developed computational theory. Care is taken to logically develop the equations applied in the analysis and clearly state all of the assumptions behind their derivation. The conclusions of the studies presented are often explored further in a theoretical manner to elucidate additional information.

The results presented in this dissertation differ substantially from the understanding of the GroEL/GroES system that exists throughout the current literature. As such, they provide a basis for future studies that will help usher in a new understanding of the chaperonin system. The updated view will likely not be as purely concerted and deterministic as the current outlook, but

rather will emphasize the importance of random and asymmetric processes in the GroEL/GroES catalytic cycle.

6.2 REFERENCES

1. Blackman, M, J Corrie, J Croney, G Kelly, J Eccleston and D Jameson. "Structural and Biochemical Characterization of Fluorogenic Rhodamine Labeled Malarial Protease Substrate." *Biochemistry* 41, (2002).

2. Hamman, B. D., A. V. Oleinikov, G. G. Jokhadze, D. E. Bochkariov, R. R. Traut and D. M. Jameson. "Tetramethylrhodamine Dimer Formation as a Spectroscopic Probe of the Conformation of Escherichia Coli Ribosomal Protein L7/L12 Dimers." *J Biol Chem* 271, no. 13 (1996): 7568-73.

3. Okoh, M. P., J. L. Hunter, J. E. Corrie and M. R. Webb. "A Biosensor for Inorganic Phosphate Using a Rhodamine-Labeled Phosphate Binding Protein." *Biochemistry* 45, no. 49 (2006): 14764-71.

4. Grason, J. P., J. S. Gresham and G. H. Lorimer. "Setting the Chaperonin Timer: A Two-Stroke, Two-Speed, Protein Machine." *Proc Natl Acad Sci U S A* 105, no. 45 (2008): 17339-44.

5. Grason, J. P., J. S. Gresham, L. Widjaja, S. C. Wehri and G. H. Lorimer. "Setting the Chaperonin Timer: The Effects of K+ and Substrate Protein on Atp Hydrolysis." *Proceedings of the National Academy of Sciences* 105, no. 45 (2008): 17334-17338.

6. Sameshima, T., R. Iizuka, T. Ueno and T. Funatsu. "Denatured Proteins Facilitate the Formation of the Football-Shaped Groel-(Groes)2 Complex." *Biochem J* 427, no. 2 (2010): 247-54.

7. Yang, D., X. Ye and G. H. Lorimer. "Symmetric Groel:Groes2 Complexes Are the Protein-Folding Functional Form of the Chaperonin Nanomachine." *Proc Natl Acad Sci U S A* 110, no. 46 (2013): E4298-305.

8. Ye, X. and G. H. Lorimer. "Substrate Protein Switches Groe Chaperonins from Asymmetric to Symmetric Cycling by Catalyzing Nucleotide Exchange." *Proc Natl Acad Sci U S A* 110, no. 46 (2013): E4289-97.

9. Madan, D., Z. Lin and H. S. Rye. "Triggering Protein Folding within the Groel-Groes Complex." *Journal of Biological Chemistry* 283, no. 46 (2008): 32003-32013.

10. Motojima, Charu Chaudhry, W. A. Fenton, G. W. Farr and A. L. Horwich. "Substrate Polypeptide Presents a Load on the Apical Domains of the Chaperonin Groel." *PNAS* 101, no. 42 (2002): 15005.

11. Todd, M. J., G. H. Lorimer and D. Thirumalai. "Chaperonin-Facilitated Protein Folding: Optimization of Rate and Yield by an Iterative Annealing Mechanism." *Proc Natl Acad Sci U S A* 93, no. 9 (1996): 4030-5.

12. Horwich, Arthur L., Adrian C. Apetri and Wayne A. Fenton. "The Groel/Groes Cis Cavity as a Passive Anti-Aggregation Device." *FEBS Letters* 583, no. 16 (2009): 2654-2662.

13. Lin, Zong, Damian Madan and Hays S. Rye. "Groel Stimulates Protein Folding through Forced Unfolding." *Nature Structural; Molecular Biology* 15, no. 3 (2008): 303-311.

14. Frank, G. A., M. Goomanovsky, A. Davidi, G. Ziv, A. Horovitz and G. Haran. "Out-of-Equilibrium Conformational Cycling of Groel under Saturating Atp Concentrations." *Proc Natl Acad Sci U S A* 107, no. 14 (2010): 6270-4.

15. Ma, J., P. B. Sigler, Z. Xu and M. Karplus. "A Dynamic Model for the Allosteric Mechanism of Groel." *J Mol Biol* 302, no. 2 (2000): 303-13.

16. Papo, N., Y. Kipnis, G. Haran and A. Horovitz. "Concerted Release of Substrate Domains from Groel by Atp Is Demonstrated with Fret." *J Mol Biol* 380, no. 4 (2008): 717-25.

17. Fei, X., D. Yang, N. LaRonde-LeBlanc and G. H. Lorimer. "Crystal Structure of a Groel-Adp Complex in the Relaxed Allosteric State at 2.7 a Resolution." *Proc Natl Acad Sci U S A* 110, no. 32 (2013): E2958-66.

18. Gorovits, B. M., J. Ybarra and P. M. Horowitz. "Atp Hydrolysis Is Critical for Induction of Conformational Changes in Groel That Expose Hydrophobic Surfaces." *J Biol Chem* 272, no. 11 (1997): 6842-5.

19. Ranson, N. A., N. J. Dunster, S. G. Burston and A. R. Clarke. "Chaperonins Can Catalyse the Reversal of Early Aggregation Steps When a Protein Misfolds." *J Mol Biol* 250, no. 5 (1995): 581-6.

APPENDIX A.1

A.1.1 Instrumentation

Absorbance measurements were carried out on a Cary 100 Bio UV Spectrophotometer. Steady state fluorescence was monitored using a Perkin Elmer LS50B Luminescence Spectrophotometer. Stop flow measurements were carried out using Applied Photophysics SX.18MV-R Stopped Flow Reaction Analyzer. HPLC analysis was performed on a Varian Pro Star Series HPLC.

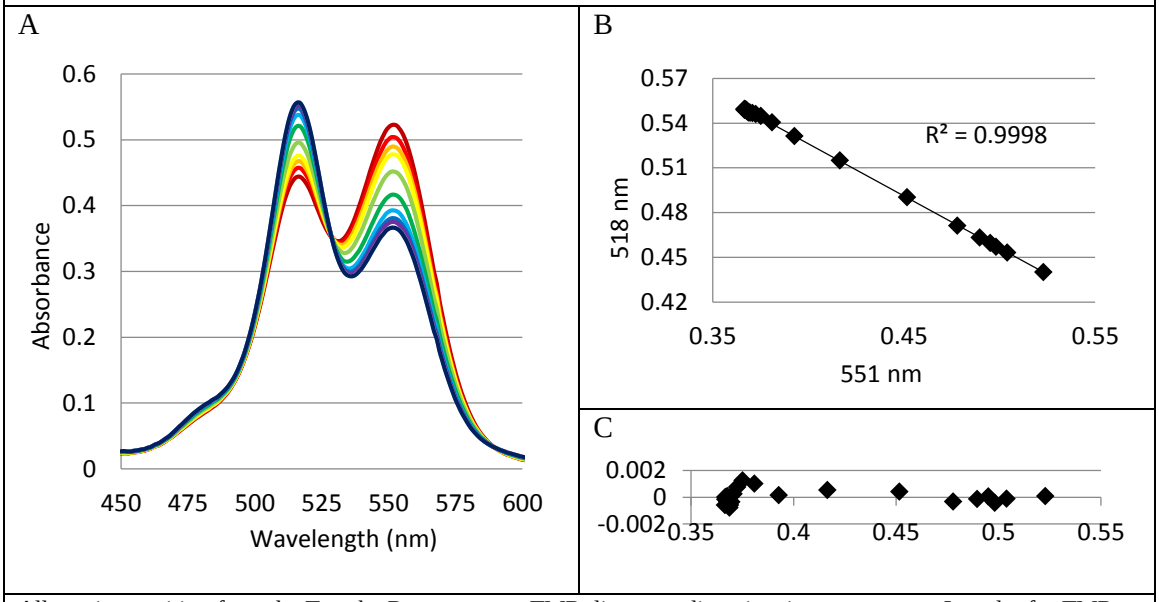
A.1.2 Labeling GroEL K242C

Numerous samples of TMR labeled EL242C were prepared by mixing substochiometric amounts of TMR into samples of EL242C in 50mM Tris pH 7.8, 2mM EDTA, 10mM NaCl. Labeling reactions were quenched with 10mM DTT after 2-4 hours and samples were passed through separate GE PD-10s to remove the remaining free dye. Samples were stored in the same buffer at 4°C for short term storage and -80°C, after being flash frozen in methanol, for long term storage.

The concentration and extent of labeling in each sample was calculated by measuring the absorbance spectra in 6M guanidine hydrochloride. The extinction coefficient of TMR-5-maleimide in 6M GHCl was experimentally determined to be 93,000 ± 1,000 $cm^{-1}M^{-1}$. The various labeled samples are prepared in such a way that protein concentration is kept constant. Plotting the absorbance at 551nm verse 280nm gives yields a straight line, with y-intercept proportional to the total protein concentration in the sample, avoiding the error incurred when trying to factor out the molar absorptivity of the dye at 280nm.

A.1.3 Monomers in T State and R state are Spectroscopically Indistinguishable: Support for the Two Species Model

Figure A.1.1 – Two Species Model and the Allosteric Cycle



Allosteric transition from the T to the R state causes TMR dimers to dissociate into monomers. In order for TMR to be a suitable allosteric probe, its spectral properties cannot change when the enzyme undergoes an allosteric transition. (A) A series of absorbance spectra, of a sample of EL242C-TMR with 5.9 TMR per ring, taken over 30 minutes. At t=0, 300µM ATP is added to the sample in the presence of 1mM K^+, shifting the allosteric equilibrium in favor of the R state (dark red). As time elapses, ATP is converted to ADP and the allosteric equilibrium is shifted back towards the T state (purple). Similar to the analysis performed in Figure 4.B and Figure 4.C, the spectra are examined to ensure that only two species are present. A plot of the absorbance at 551nm vs 518nm for each curve. (C) The residuals calculated from fitting the plot with a line. Although the residuals are not completely random, the fit is good, R^2=0.9998.

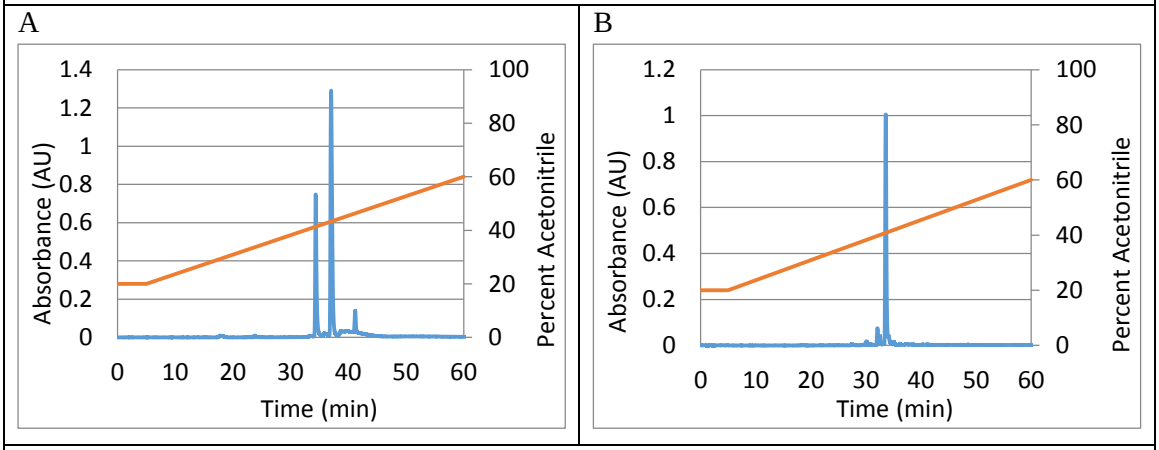
A.1.4 Dithiotheitol-tetramethylrhodamine (TMR_2-DTT) as a model system

Stock solutions of TMR-5-maleimide (Invitrogen) were prepared in DMSO. The concentration was measured using the extinction coefficient 95,000 $cm^{-1}M^{-1}$ at 541nm. A model TMR dimer sample, TMR_2-DTT, was prepared by mixing a 1:2 ratio of DTT to TMR-5-maleimide in buffer containing 10mM Tris pH 7.5 and 1mM EDTA. TMR-5-maleimide alone was taken to be the model monomer sample. TMR_2-DTT was further purified on an Agilent Semi Prep Eclipse XDB-C18 column (5µm pores, 9.4 x 250mm, 1mL/min flow rate) using an H_2O: acetonitrile gradient (Figure SI.2A). SDS, a surfactant that disrupts the hydrophobic stacking interaction in TMR dimers, was used to calculate the population of total dye molecules in a sample. After the addition of SDS, both monomer and dimer samples have the same absorbance spectra. Two peaks of TMR_2-DTT elute off the column, both having similar absorbance and fluorescence characteristics.

The two dimer peaks consist of predominantly stacked populations. Nonetheless, the dimer samples are not exactly identical. Both the fluorescence and absorbance properties of the two samples indicate that the first peak has a larger population in the non-stacked form. Furthermore, the fact that their retention on the C18 column differs, indicates a structural difference. On the other hand ESI-MS analysis of each peak yielded the same spectrum. We suggest that the emergence of two peaks is a result of the new chiral center created when the maleimide group reacts with the thiol. This produce three distinct populations, the R'-R', S'-S' and, the R'-S'. In the parallel arrangement thought to be assumed by the dimers, the xanthene rings are thought to be aligned plane parallel to one another. The R'-R' and S'-S' dimers are enantiomers, which are distinct from those available to the R'-S' dimers. Thus, despite the presence of three distinct populations, only two distinct peaks were observed.

The chirality explanation for the two peaks is supported by two pieces of evidence. First, both peaks produce identical mass spectra that have the predicted m/z and isotopomer distribution for a molecule of TMR_2-DTT. Secondly, TMR_2-DTT dimers prepared with TMR-iodoacetamide do not form a chiral center when reacted with DTT. When samples of $T5I_2$-DTT were purified on the C18 column, using the same protocol as TMR-5-maleimide dimers, only a single dimer peak eluted

Figure A.1.2 – HPLC Purification of TMR_2-DTT Dimers

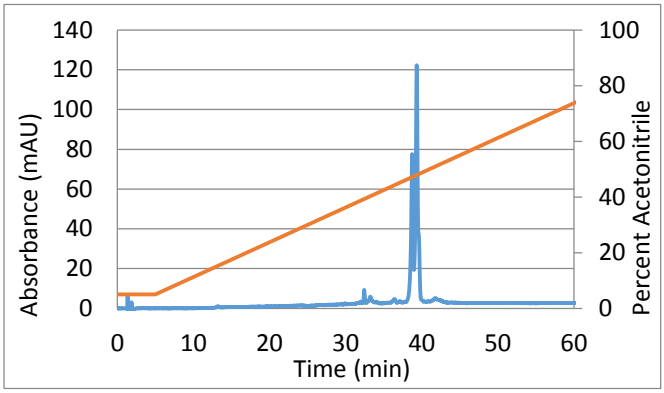


Chromatagram of the absorbance at 551nm (blue) and the applied acetonitrile gradient (red, secondary axis). (A) Purification of TMR-5-maleimide dimers. (B) Purification of TMR-5-iodoacetamide dimers.

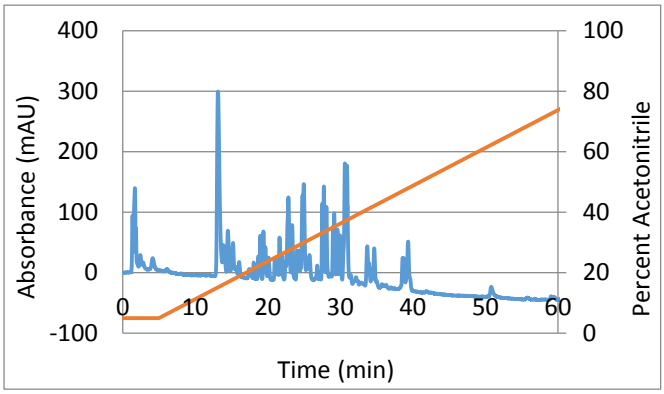
from the column (Figure SI.2B).

Figure A.1.3 – HPLC Analysis of EL242C-TMR Tryptic Digest

A



B



Sample of EL242C-TMR with an average of 5.6 TMR/ring was treated with trypsin and applied to a C18 column. Chromatagram of the absorbance (blue) and the applied acetonitrile gradient (red, secondary axis). (A) Absorbance monitored at 551nm, showing the elution of TMR. (B) Absorbance monitored at 215nm, showing the elution of peptide fragments.

A.1.5 Specificity of TMR labeling: Tryptic digests

The proposed stochastic binding model, and the subsequent application of the binomial distribution to the analysis of the EL242C-TMR system, relies on the assumption that only the modified 242C residues are being labeled. Although GroEL has three native cysteines per subunit in its wild type sequence, they are all relatively well buried compared to residue 242, lying deep within the barrel-like structure of GroEL. The assumption is further supported by the tryptic digest analysis of EL242C-TMR samples. This was performed using an Agilent Eclipse XDB-C18 column (5µm pore, 2.1 x 150mm, 0.25 mL/min flow rate). After incubating with trypsin, samples of EL242C-TMR analyzed on a C18 column produced two dominant peaks with similar retention times (Figure SI.3). The two peaks are thought to correspond to the two stereochemically distinct species that occur when using a maleimide linker to label with TMR.

A.1.6 Singular Value Decomposition

Singular value decomposition (SVD) was used in a principle component analysis of the spectra of EL242C-TMR to support the two species model proposed. SVD was performed in MATLAB using absorbance spectra measured in the Cary 100 Bio UV Spectrophotometer. In both experiments, over 98% of the variance is accounted for using the first two singular values.

A.1.6 Fluorescence verse ATP

Experiments were performed in the stop flow apparatus. The wavelength of excitation in these experiments was 540nm and a 575nm cutoff filter was used. Both entrance and exit slit width of the monochromator was 4.65nm and the PMT set to 470V. Experiments were carried out at 36°C in 50mM Tris pH7.5, 10mM KCl, 10mM MgCl, 90mM NaCl, 2mM PEP, 5U of PK and 8U of LD. Although the stop flow produces data that is a function of time, only the steady state portion of the data sets was used in this analysis. Error reported is the standard deviation of repeated measurements on the instrument under similar conditions.

APPENDIX A.2

A.2.1 Background of Pre-Steady State Kinetic Analysis

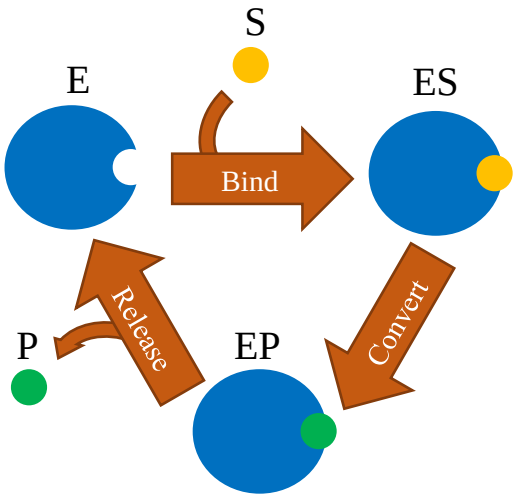
An accurate model for the catalytic mechanism of an enzyme is required not only to describe its functional behavior in vitro, but also within reaction networks in vivo. These mechanisms consist not only of traditional chemical reaction mechanisms (bond making and breaking), but also other thermodynamically governed processes commonly exhibited by enzymes, such as component association and local structural rearrangements. Rapid mixing experiments are valuable ways to experimentally test proposed catalytic mechanisms because they allow for the potential observation of species that may be only sparsely populated in the steady state.

The catalytic mechanism of an enzyme can be represented by a **kinetic scheme**: a network of distinct states that are connected by the various processes that occur in enzyme catalysis. An example of the catalytic mechanism and kinetic scheme of a hypothetical enzymatically catalyzed reaction, as well as its corresponding free energy diagram, are given in Figure A.2.1. The states in a kinetic scheme represent any of the distinct species present throughout the reaction and are representative of the reaction coordinate corresponding to the local minima of the free energy diagram of the reaction. The processes linking these states are the thermodynamically governed processes discussed above and are representative of the free energy barriers that separate the various states in a kinetic scheme.

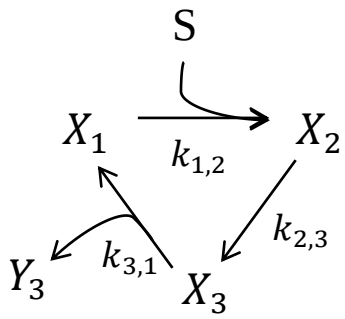
The reason for conversion of a catalytic mechanism to a kinetic scheme is because it allows for its quantitative analysis. Kinetic schemes have a corresponding mathematical characterization as a series of coupled first-order differential equations, sometimes referred to as the chemical master equation. The states of the kinetic scheme comprise the set of time dependent state functions which correspond to the population of each species at time t. Throughout this text, the states corresponding to any of the enzyme species will be represented by indexed functions $X_i(t)$. The

Figure A.2.1 – Relationship Between Catalytic Mechanism, Kinetic Scheme and the Free Energy Diagram of an Enzyme Catalyzed Reaction

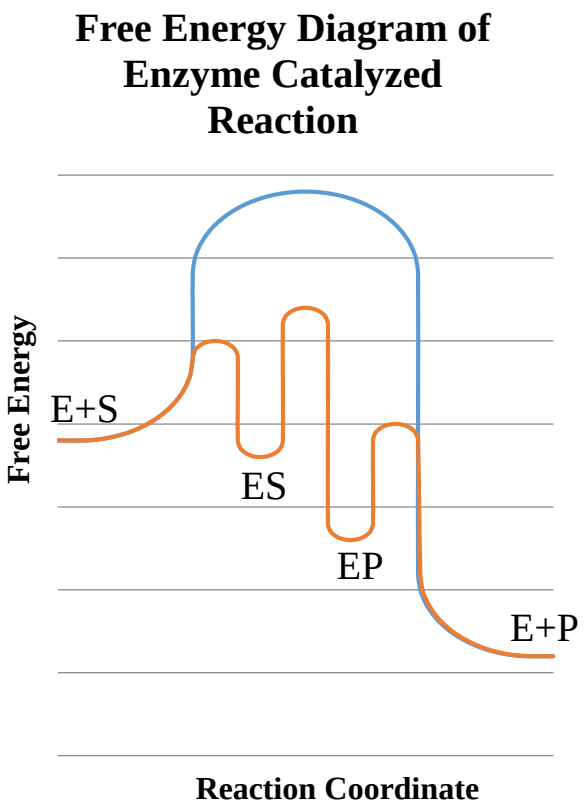
A



B



C



(A) Catalytic Model or Catalytic System of a hypotentical enzyme. The model describes all species present in solution throughout a enzymatically catalyzed reaction (colored shapes - E, ES, EP, P, S) and the various processes that link them (red arrows – substrate binding, conversion from substrate to product, release of product). (B) Kinetic Scheme is the conversion of a catalytic mechanism to a mathematical model that is experimentally testable. The conversion established a quantitative link between the species in the catalytic model. The various species in the catalytic mechanism represent states in the kinetic scheme, X_i, and the processes connecting the states are represented by micro-constants $k_{i,j}$. The kinetic scheme presented is assumed to be used under pseudo first order conditions described in the text. (C) Free Energy Diagram of the hypothetical enzyme catalyzed reaction modeled in (A). The rate parameters in the kinetic scheme (B) can be related to the heights of the free energy barriers in the free energy diagram (C) through the Arrhenius equation.

Table A.2.1 – Conversion from Catalytic Mechanism to the Chemical Master Equation of an Enzymatically Catalyzed Reaction System		
Species	State Function	Term in Chemical Master Equation
E	$X_1(t)$	$\frac{dX_1}{dt} = -Sk_{1,2}X_1 + k_{3,1}X_3$
ES	$X_2(t)$	$\frac{dX_2}{dt} = -k_{2,3}X_2 + Sk_{1,2}X_1$
EP	$X_3(t)$	$\frac{dX_3}{dt} = -k_{3,1}X_3 + k_{2,3}X_2$
P	Not State but Flux Function $Y_3(t)$	$\frac{dY_3}{dt} = k_{3,1}X_3$

subset of states $\{X_i(t)\}$ is considered a conserved set of states, due to the fact that their sum is constant throughout the experiment. All other states will be represented by indexed functions $Y_i(t)$, and it will be shown that such states can be obtained from knowledge of the set of functions $\{X_i(t)\}$.

The processes linking the various states of the system are mathematically represented as the set of microscopic rate constants, $\{k_{i,j}\}$, that correspond to the rate of the process that feeds from state *i* into state *j*. The set of microscopic rate constants, $\{k_{i,j}\}$, will be referred to throughout the text as **micro-constants** to distinguish them from the other rate terms found in equations used throughout the analysis. This concept fits into broader chemical theory where the values of these micro-constant rates are related back to the free energy diagram of the process as barrier heights through the Arrhenius equation and Transition State Theory.

The state functions and micro-constants in the kinetic scheme can be organized into a series of coupled first-order differential equations, making up its chemical master equation, Table A.2.1. These equations give each state function's rate of change (derivative in time) relative to the various micro-constants and one or more of the system's state functions. The rate of change of state is equal to the sum of processes feeding into the state, minus the sum of processes out of the

state. The system of equations has an exact solution for a given set of initial conditions $\{X_i(0)\}$, ie the concentration of each species at $t = 0$.

In rapid mixing experiments, the population of one or more states is monitored as a function of time over the course of the experiment. The experiments are usually carried out under pseudo-first order conditions: the concentration of substrate is much greater than the concentration of total enzyme and the length of the experiment is very short compared to the decay of the substrate population. This allows one to assume the state function corresponding to the substrate species is constant in time and linearizes the system of first-order ODEs governing the behavior of the system.

The solution to the given system of equations can be obtained for a given set of initial conditions. Looking again at the system's chemical master equation, the state function corresponding to the product, $Y_3(t)$, has a special form that is solely a function of $X_3(t)$ alone. This allows us to solve for $Y_3(t)$ by integration. More generally, any function of the form $Y_i(t)$ in Eq A.2.1 is a measure of the probability flux out of state *i*. Functions of this form will be referred to throughout the text as **flux functions**. Product accumulation data thus has the functional form of a flux function. These functions can be obtained by first finding the solution to the set of equations linking the states $\{X_i(t)\}$ of the system and then integrating the relevant state functions.

$$Y_i(t) \equiv \int_0^t k_i X_i(s)\, ds \qquad \text{Eq A.2.1}$$

The master equation governing the coupled states of a system can be represented as a series of first order differential equations, Table A.2.1. They can be more compactly arranged as a differential matrix equation, Eq A.2.2, where the states of the system, $\{X_i(t)\}$, are arranged into a state vector, $\vec{X}(t)$, and the rates of the processes relating the states are arranged into a rate matrix, $\widehat{M}$. If kij is the micro-constant governing the process by which state i is converted to state j, then

the value of the matrix element $\widehat{M}(i,j) = k_{i,j}$. The diagonal elements of the rate matrix are the negative of the sum of the all rates leading out of the state, ie

$$\widehat{M}(i,i) = -\sum_{j \neq i}^{N} k_{i,j}$$

The differential matrix equation formulation corresponding to the system in Figure A.2.1 and the chemical master equation of Table A.2.1 is given below. Note, the flux function corresponding to the product accumulation has been omitted from the equation, since it can be obtained from the solution to Eq A.2.2.

$$\frac{d\vec{X}}{dt} = \widehat{M}\vec{X} \qquad \text{Eq A.2.2}$$

$$\widehat{M} = \begin{bmatrix} -Sk_{1,2} & 0 & k_{3,1} \\ Sk_{1,2} & -k_{2,3} & 0 \\ 0 & k_{2,3} & -k_{3,1} \end{bmatrix} \qquad \vec{X}(t) = \begin{bmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{bmatrix}$$

A.2.2 General Solution

In a system with N conserved states the rate matrix will be an NxN square matrix. The differential matrix equation can be solved to obtain the general form:

$$\vec{X}(t) = \sum_{i=1}^{N} c_i \, e^{\lambda_i t} \, \vec{u}_i \qquad \text{Eq A.2.3}$$

Where $\{\lambda_i\}_{i=1}^{N}$ is the set of N eigenvalues of the rate matrix $\widehat{M}$, $\{\vec{u}_i\}_{i=1}^{N}$ is its corresponding set of eigenvectors and $\{c_i\}_{i=1}^{N}$ is a set of constants that are determined by the initial conditions of the system. Thus, each state function of the system will have the general form

$$X_i(t) = \sum_{j=1}^{N} A_{i,j} \, e^{\lambda_j t} \qquad \text{Eq A.2.4}$$

In a system of N states, the function corresponding to the population of any species, or any combination of species, is thus a function of N exponential terms. The rate constants of the state functions, $\{\lambda_i\}_{i=1}^{N}$ (not to be confused with the micro-constants of the system), are the eigenvalues of $\widehat{M}$ and can be found by solving the characteristic polynomial of $\widehat{M}$, obtained by

$$\det(\widehat{M} - \lambda \hat{I}) = 0$$

where $\hat{I}$ is the NxN identity matrix. The zeroth order term in λ of the characteristic polynomial is always simply $\det(\widehat{M})$. In a rate matrix of the kind outlined above, the columns of the rate matrix sum to zero, indicating the linear dependence of the row vectors. It follows that $\det(\widehat{M})$ is equal to zero in such systems. This means that one of the solutions to the characteristic polynomial of matrix $\widehat{M}$ will always be zero. Plugging this result into Eq A.2.4, the functional form of each state will thus contain a constant term and the sum of N-1 exponential terms:

$$X_i(t) = A_1 + \sum_{j=2}^{N} A_{i,j}\, e^{\lambda_j s} \qquad \text{Eq A.2.5}$$

Similarly, we can utilize the new form of the general solution of a state function to obtain the general form of the flux function $Y_i(t)$, by plugging Eq A.2.5 into Eq A.2.1:

$$Y_i(t) = k_i \int_0^t A_1 + \sum_{j=2}^{N} A_{i,j}\, e^{\lambda_j s}\, ds$$

$$Y_i(t) = k_i \left[A_{i,1} t + \sum_{j=2}^{N} \frac{A_{i,j}}{\lambda_j}\, e^{\lambda_j t} - \sum_{j=2}^{N} \frac{A_{i,j}}{\lambda_j}\, e^{\lambda_j 0} \right]$$

$$Y_i(t) = k_i \left[A_{i,1} t + \sum_{j=2}^{N} \frac{A_{i,j}}{\lambda_j}\, e^{\lambda_j t} - \sum_{j=2}^{N} \frac{A_{i,j}}{\lambda_j} \right]$$

Combining all the constant coefficient in each term into new constants, we obtain the general form of the flux function Y_i of state X_i:

$$Y_i(t) = S * t + K + \sum_{j=2}^{N} B_{i,j}\, e^{\lambda_j t} \qquad \text{Eq A.2.6}$$

$$\lim_{t\to\infty} Y_i(t) = S * t + K \qquad \text{Eq A.2.7}$$

Taking the limit as time goes to infinite of Eq A.2.6, one obtains Eq A.2.7, is the "steady state" portion of the probability flux function. Thus, in a given system, the accumulation of product as a function of time contains N-1 exponential terms, a constant term and a term that is linear in time. If the steady state of the system is reached on the time scale of the experiment, the values for S and K can be obtained by least squares fitting of a line to the linear portion of the graph.

A.2.3 Reduction of States

Because the number of states in a kinetic scheme is equal to the number of terms in the corresponding state functions, it is often desirable to simplify the kinetic scheme of a given catalytic mechanism whenever possible. The most common simplification is to combine states in a catalytic mechanism that are linked by very fast events into a single state in the corresponding kinetic scheme. We will justify this simplification using the model system given below.

$$X_1 \overset{k_1}{\rightarrow} X_2 \overset{k_2}{\rightarrow} X_3$$

$$X_1(0) = 1;\ X_2(0) = 0;\ \ X_3(0) = 0$$

$$\frac{dX_1}{dt} = -k_1 X_1$$

$$\frac{dX_2}{dt} = -k_2 X_2 + k_1 X_1$$

$$\frac{dX_3}{dt} = k_2 X_2$$

$$X_1(t) = e^{-k_1 t}$$

$$X_2(t) = \frac{k_1}{k_1 - k_2}(e^{-k_2 t} - e^{-k_1 t})$$

$$X_3(t) = 1 - \frac{k_1}{k_1 - k_2} e^{-k_2 t} + \frac{k_2}{k_1 - k_2} e^{-k_1 t}$$

If $k_2 \gg k_1$, then

$$k_1 - k_2 \approx -k_2$$

$$\frac{k_1}{k_2} \approx 0$$

$$X_2(t) \approx \frac{k_1}{-k_2}\left(e^{-k_2 t} - e^{-k_1 t}\right) \approx 0$$

$$X_3(t) \approx 1 - \frac{k_1}{-k_2} e^{-k_2 t} + \frac{k_2}{-k_2} e^{-k_1 t} \approx 1 - e^{-k_1 t}$$

The simplicity of the system allows us to solve it exactly and obtain the analytic form of the set of state functions, $\{X_i(t)\}_{i=1}^{3}$. In the limit of large k_2, the state function $X_2(t)$ is approximately zero for all t. The state function $X_3(t)$ is reduced to a form that contains no reference to the micro-constant k_2. The simplified set of state functions corresponds to the solution of the simplified two-state kinetic scheme below, which does not contain the state X_2 at all.

$$X_1 \overset{k_1}{\rightarrow} X_3$$

$$X_1(0) = 1;\ X_3(0) = 0$$

A.2.4 Analytical Solution to a System of Coupled First Order ODEs

For a given set of initial conditions, $\vec{X}(0)$, the analytical form of the state vector is most easily obtained using the inverse Laplace transform method:

$$\vec{X}(t) = \mathcal{L}^{-1}[(s\,I - M)^{-1} * \vec{X}(0)] \qquad \text{Eq A.2.8}$$

Here, $\mathcal{L}^{-1}[\ ]$ is the inverse Laplace transform, $(\quad)^{-1}$ is the inverse matrix operation, s is the Laplace space independent variable, I is the identity matrix and, M is the rate matrix for the system.

The inverse Laplace transform method gives the analytical form of the state vector: the population of each species in a system as a function of time and its rate constants. However, in all but the smallest systems, if the rate constants of the system are not known in advanced, these analytic solutions are overly complex and difficult to use. For example, in a system of N states, the functional form of the eigenvalues alone will be the solution to an N-1 polynomial. Often, it is more useful to solve the system of ODEs numerically. MATLAB has numerous ODE solving algorithms.

A.2.5 Scaling Analysis

Scaling analysis is an analytical technique that can be used to simplify an equation by assuming a simple relationship between the magnitudes of its independent parameters. It is especially useful in analyzing symmetric functions because the variables can be ordered without loss of generality. A small scaling parameter is introduced into the equation that roughly relates the independent parameters. The function is reevaluated, assuming the terms that are highest in order with respect to the scaling parameter are negligible. Reevaluation of the equation can sometimes yield drastically simplified results.

Consider a function $F(x; a, b, c, d)$ which is symmetric in the parameters $\{a, b, c, d\}$, meaning that $F(x; a, b, c, d) = F(x; d, b, c, a) = F(x; b, c, a, d) = etc.$ for any permutation of the set

$\{a, b, c, d\}$. The set can thus be ordered into a new set $\{\alpha_i\}_{i=1}^{4}$ where $\alpha_i \geq \alpha_{i+1}$ for all i without changing the outcome of the function F.

Now, let's assume that we have reason to believe that some of the various α_i differ from one another by about an order of magnitude. Then we can introduce a small scaling parameter, ε, where $\varepsilon < 0.1$. We then adjust the set of parameters by including the scaling parameter such that:

$$\alpha_1 = \beta_1 \qquad \alpha_2 = \varepsilon\,\beta_2 \qquad \alpha_3 = \varepsilon^2\,\beta_3 \qquad \alpha_4 = \varepsilon^2\,\beta_4 \qquad \text{and} \qquad \beta_1 \approx \beta_2 \approx \beta_3 \approx \beta_4$$

To see why such an assumption may be useful, consider the case where:

$$F(x; a, b, c, d) = x^3 + (b + c + a + d)\,x^2 + (bc + ab + db + ac + ad + dc)x + adb + abc + bdc + adc$$

The function does not factor and the solution to the equation $F(x; a, b, c, d) = 0$ can only be found by using the cubic root equation. However, assumed the above detailed scaling relationship between parameters and plug back into F. $F(x; \beta_1, \varepsilon\,\beta_2, \varepsilon^2\,\beta_3, \varepsilon^2\,\beta_4)$ contains terms ranging from zero to fifth order in the scaling parameter ε. $\varepsilon < 0.1$ implies that $\varepsilon^4 < 0.0001$. Dropping from $F(x; \beta_1, \varepsilon\,\beta_2, \varepsilon^2\,\beta_3, \varepsilon^2\,\beta_4)$ the terms that are fourth and fifth order in ε, one can now factor the polynomial:

$$F\left(x; \{\alpha_i\}_{i=1}^{4}\right) = (x + \alpha_1)(x + \alpha_2)(\alpha_3 + \alpha_4 + x)$$

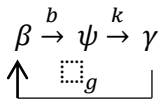
Scaling analysis can be very useful in analyzing kinetic schemes, where it is often the case that different events are occurring on very different timescales. It can also be used to justify equating a fitted parameter to a system micro-constant.

APPENDIX A.3

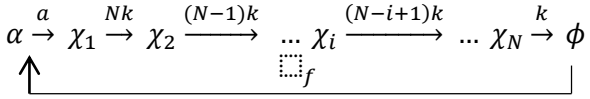
A.3.1 Proof of Claim that Stochastic and Concerted Hydrolysis Mechanisms are Indistinguishable During the First Pass

Figure A.3.1 – Starting Cyclic Kinetic Schemes

(A) – Concerted



(B) – Stochastic



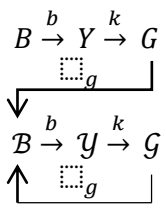
Consider the concerted and stochastic cyclic kinetic schemes presented in Figure A.3.1. Assuming the initial state of the system is that the entire population resides entirely in state α for the stochastic cycle and β for the concerted cycle. Using the equalities given in Table A.3.1, the cyclic systems of Figure A.3.1 can be converted into the kinetic schemes of Figure A.3.2. The kinetic schemes in Figure A.3.2 allow us to separate a state of the system (greek letter) into the sum of states corresponding to the first pass of occupancy (straight letters) and all other passes (fancy letters).

Table A.3.1 – Equalities Used to Decompose the Cyclic Kinetic Schemes

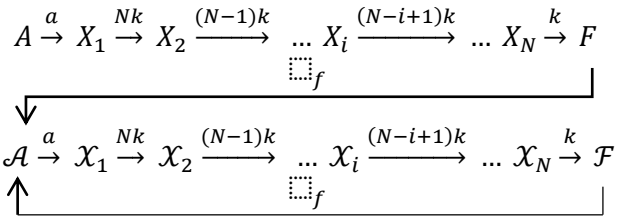
Concerted	Stochastic
$\beta = B + \mathcal{B}$	$\alpha = A + \mathcal{A}$
$\psi = Y + \mathcal{Y}$	$\chi_i = X_i + \mathcal{X}_i$
$\gamma = G + \mathcal{G}$	$\phi = F + \mathcal{F}$

Figure A.3.2 – Conversion of Cyclic Kinetic Scheme to a Combined Linear and Cyclic Kinetic Scheme

(A) – Concerted



(B) – Stochastic



The set of coupled first order ODEs can be obtained by inspection of the kinetic scheme in Figure A.3.2 and are given in Table A.3.2 for the concerted system and Table A.3.3 for the stochastic system.

Table A.3.2 – System of Coupled ODE Equations for the Concerted System

First Pass	Subsequent Passes
$\dot{B} = -b\,B$	$\dot{\mathcal{B}} = -b\mathcal{B} + gG + g\mathcal{G}$
$\dot{Y} = -kY + b\,B$	$\dot{\mathcal{Y}} = -k\mathcal{Y} + b\,\mathcal{B}$
$\dot{G} = -gG + kY$	$\dot{\mathcal{G}} = -g\mathcal{G} + k\mathcal{Y}$

Note: Time derivatives are represented in dot notation.

Table A.3.3 – System of Coupled ODE Equations for the Stochastic System

First Pass	Subsequent Passes
$\dot{A} = -a\,A$	$\dot{\mathcal{A}} = -a\,\mathcal{A} + fF + f\mathcal{F}$
$\dot{X}_1 = -NkX_1 + a\,A$	$\dot{\mathcal{X}}_1 = -Nk\mathcal{X}_1 + a\,\mathcal{A}$
$\dot{X}_2 = -(N-1)kX_2 + NkX_1$	$\dot{\mathcal{X}}_2 = -(N-1)k\mathcal{X}_2 + Nk\mathcal{X}_1$
$\dot{X}_i = -(N-i+1)kX_i + NkX_{i-1}$	$\dot{\mathcal{X}}_i = -(N-i+1)k\mathcal{X}_i + Nk\mathcal{X}_{i-1}$
$\dot{F} = -fF + kX_N$	$\dot{\mathcal{F}} = -f\mathcal{F} + k\mathcal{X}_N$

Note: Time derivatives are represented in dot notation.

Now the product accumulation functions for the two sytems can also be reformulated in terms of the first pass and subsequent pass state functions, Table A.3.4.

Table A.3.4 – Product Accumulation Functions

Concerted	Stochastic
$\Phi = \Phi_I + \Phi_{II}$	$\Psi = \Psi_I + \Psi_{II}$
$\Phi_I = \int_0^t k\, Y\, ds$	$\Psi_I = \int_0^t \sum_{i=1}^{N} (N-i+1) k\, X_i\, ds$
$\Phi_{II} = \int_0^t k\, \mathcal{Y}\, ds$	$\Psi_{II} = \int_0^t \sum_{i=1}^{N} (N-i+1) k\, \mathcal{X}_i\, ds$

We wish to show that Φ_I and Ψ_I are related to one another by a multiplicative constant. This implies that we hope to be able to reduce Ψ_I from a sum of states to a single state. Let Eq A.3.1 define a new function Z.

$$Z \equiv \sum_{i=1}^{N} (N-i+1)\, X_i \qquad \text{Eq A.3.1}$$

Redefine the state Y in the concerted system in terms of its time derivative function, Eq A.3.2. Repeat the procedure for the various states of the concerted system, Eq A.3.3 to Eq A.3.5

$$k\, Y = -\dot{Y} + b\, B \qquad \text{Eq A.3.2}$$

$$NkX_1 = -\dot{X}_1 + a\, A \qquad \text{Eq A.3.3}$$

$$(N-1)kX_2 = -\dot{X}_2 + NkX_1 \qquad \text{Eq A.3.4}$$

$$\begin{aligned} &= -\dot{X}_2 - \dot{X}_1 + a\, A \\ (N-i+1)kX_i &= -\sum_{j=1}^{i} \dot{X}_j + a\, A = k\, Z \end{aligned} \qquad \text{Eq A.3.5}$$

Plugging the results of Eq A.3.5 back into the formula for Ψ_I from Table A.3.4 gives Eq A.3.6.

$$\Psi_I = -\int_0^t \sum_{i=1}^{N} \left[-\sum_{j=1}^{i} \dot{X}_j + aA \right] ds = \int_0^t -\sum_{i=1}^{N} (N - i + 1)\, \dot{X}_j + NaA \; ds$$

Eq A.3.6

Notice:

$$\sum_{i=1}^{N} (N - i + 1)\, \dot{X}_j = \dot{Z}$$

Thus:

$$\Psi_I = \int_0^t k\, Z \; ds = \int_0^t -\dot{Z} + NaA \; ds$$

And:

$$\Phi_I = \int_0^t k\, Y \; ds = \int_0^t -\dot{Y} + b\, B \; ds$$

Thus if $a = b$ and $A(0) = B(0) = 1$:

$$\Psi_I = N\, \Phi_I$$

This is the expected relationship between a stochastic and concerted process in a linear system and indicate that the two formulations are exactly identical for the first pass of the system. It should be noted that $\Psi_{II} \neq N\, \Phi_{II}$ indicating that the steady state value of the two systems will differ.

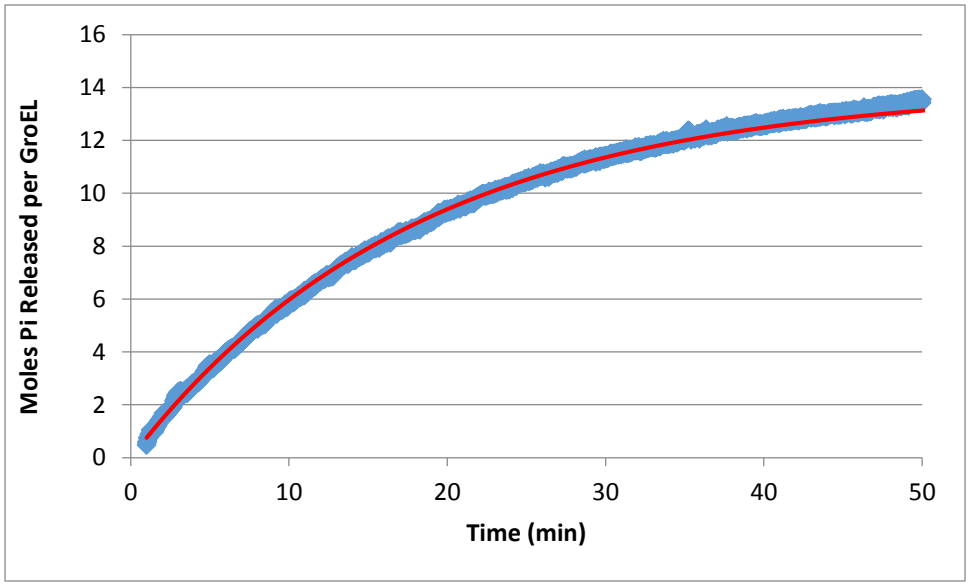
APPENDIX A.4

A.4.1 Rate of ATP Hydrolysis in the D398A Mutant

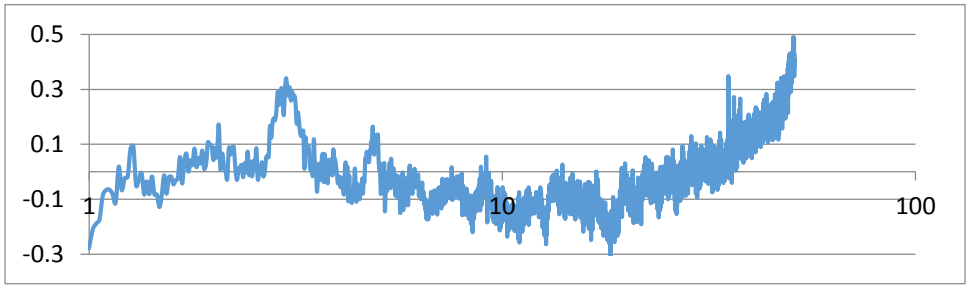
The D398A variant of GroEL hydrolyzes ATP at a rate ~ 0.1% of the wild-type protein. The rate constant of hydrolysis was obtained by fitting the Pi accumulation curve of the mutant to Eq A.4.1. The results are presented in Figure A.4.1.

Figure A.4.1 – Hydrolysis Kinetics of D398A

A



B



(A) Raw Pi versus time data in blue plotted along with the best fit exponential association in red. (B) Residuals of the fit.

$$F(t;H) = N - Ne^{-Ht} \qquad \text{Eq A.4.1}$$

The best fit value obtained is $H = 0.05568 \pm 4 \times 10^{-5}$ min^{-1} when N is fixed at 14. This is the value used throughout Chapter 4.

A.4.2 Subunit Mixing Protocol

Protocol adapted from ref (2,6,7)

1.7mL of 6M Urea combined with 300uL of 100uM EL subunits in 50mM Tris-HCl pH 7.5. Allowed to sit at room temperature for 90 minutes.

Just before incubation period is up, 40mL of refolding buffer (50mM Tris-HCl pH 7.5, 5mM $MgCl_2$, 400mM $(NH_4)_2SO_4$) is set aside and combined with ATP to 5mM. 38mL of this mixture added to EL + Urea sample, diluting the urea to 0.25M. Allowed to sit for 30 minutes before concentration.

15mL YM-50 centricon used for concentration. 40mL of sample concentrated in 3 runs of 40 minutes spinning at 3000rpm at 17C. Remaining sample sat at 4C while 15mL portion was being concentrated.

After all 40mL of sample had been concentrated to ~1mL, the remaining 1mL of sample was washed with the 14mL of the remaining refolding buffer (without ATP). This was again concentrated to ~1mL by centrifuging at 3000rpm for 40min at 17C.

There were then two washes with 50:10 buffer (50mM Tris-HCl pH 7.5, 10mM $MgCl_2$). 14mL of 50:10 buffer was added to remaining 1mL of sample. This was concentrated to ~1mL by centrifuging at 3000rpm for 40min at 17C. Repeated a second time.

A.4.3 Subunit Mixing of Wild-Type and TMR Labeled K242C

Wild type GroEL was mixed with TMR labeled K242C (introduced in Chapter 1) according to the protocol in Section A.4.2, in order to determine the effectiveness of the procedure at

producing a mixture of subunits. The mixing of TMR dimer labeled subunits with wild-type subunits produces a mixture with an average number of dye molecules per ring that is less than the original value for the labeled subunits by a factor directly proportional to the mole fraction of wild-type subunits in the mixture. If there was no mixing of subunits, the proportion of dimers to monomers would remain the same in the mixture as in the original labeled sample. If a absorbance spectrum of the sample was taken under native conditions, and normalized with respect to the total concentration of dye within the sample, samples with no mixing would overlay the original spectrum of the EL242C-TMR in the mixture. Only if subunits were exchanging would there be a decrease in the dimer population and a corresponding increase in the monomer population.

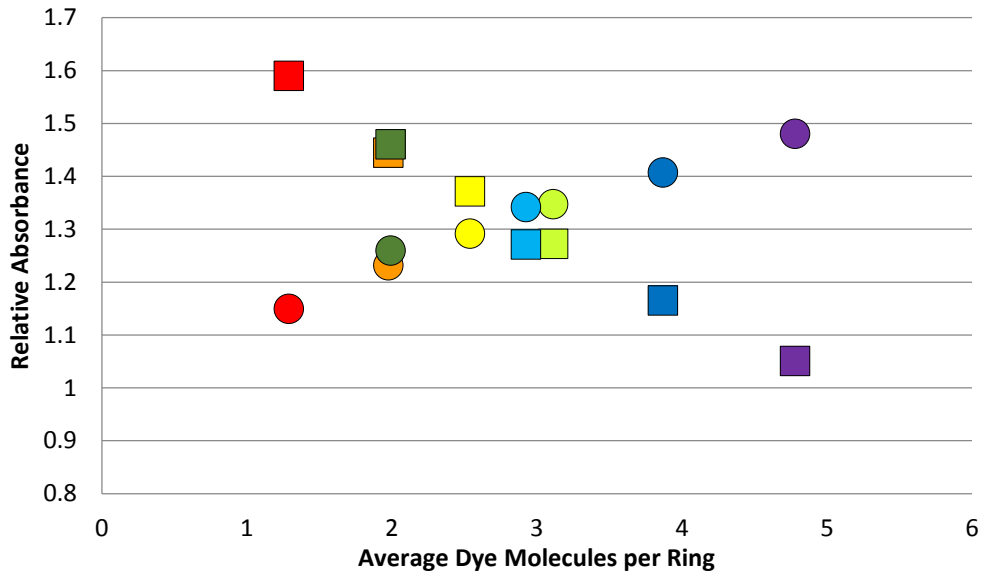
Two starting samples of EL242C-TMR were used containing 3.6 and 5.2 dye molecules per ring. Each EL242C-TMR sample was mixed with wild-type GroEL of the same concentration in the following 4 ratios: 0% WT, 20% WT, 40% WT and, 60%WT.

Table A.4.1 – Color Coding of Samples in Figure A.4.2	
Initial TMR/Ring	
3.6	5.2
242C/WT	242C/WT
100/0	**100/0**
80/20	**80/20**
60/40	**60/40**
40/60	**40/60**

The absorbance spectra of the reconstituted 14mers were recorded in the presence of GHCl in order to determine the average number of dye molecules per ring and the population of monomers and dimers in the sample. The absorption at 559nm gives the concentration of dye in the sample; the value of the extinction coefficient used was 93400 $M^{-1}cm^{-1}$. The absorbance at 280nm can be used to determine the concentration of GroEL subunits, after correcting for the contribution of the dye. As in Chapter 1, the contribution of the dye was corrected for by determining the y-intercept

of a plot of A_{551} against A_{280}. The linear fit to the mixed subunit samples has a relatively low coefficient of determination ($R^2 = 0.95$) due to slight differences in yield between samples. The extinction coefficient for GroEL used was 9600 $M^{-1}cm^{-1}$.

Figure A.4.2 – Illustration of Subunit Mixing



Absorbance of subunit mixed samples, normalized with respect to dye concentration, at 551nm (squares) and 518nm (circles). The absorbance at the respective wavelength is plotted against the calculated number of dye molecules per ring. The color coding is given in Table A.4.1. The plot demonstrates that the GroEL 14mers dissociate and reform into mixed rings: rings containing a mixture of EL242C and wild-type subunits.

The measurements were performed in the absence of GHCl, where the absorbance at 518 nm and 551 nm is directly proportional to the population of dimers and monomers, respectively. The absorbance at these two wavelengths can be calculated relative to the total number of dye molecules in the sample by dividing by the value of absorbance at isobestic point of the system, 528nm. A plot of the measured number of dye molecules per ring against the absorbance at the indicated wavelength is given in Figure A.4.2.

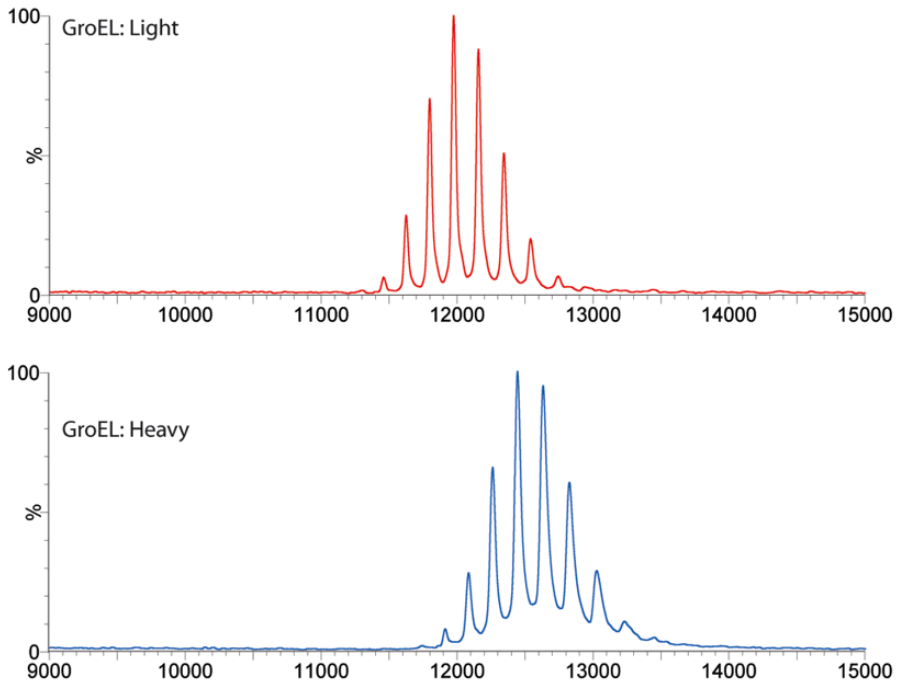
The results of the analysis in Figure A.4.2 demonstrate that the mixing of subunits by the procedure outlined in Section A.4.2 produces a mixture of subunits within the population of newly reconstituted 14mers. This is because the monomer and dimer populations, proportional to

the absorbance at 551nm and 518nm, respectively, change with the relative proportion of wild-type GroEL in the mixed sample.

A.4.4 Subunit Mixing of Heavy and Light GroEL Monitored by Mass Spectroscopy

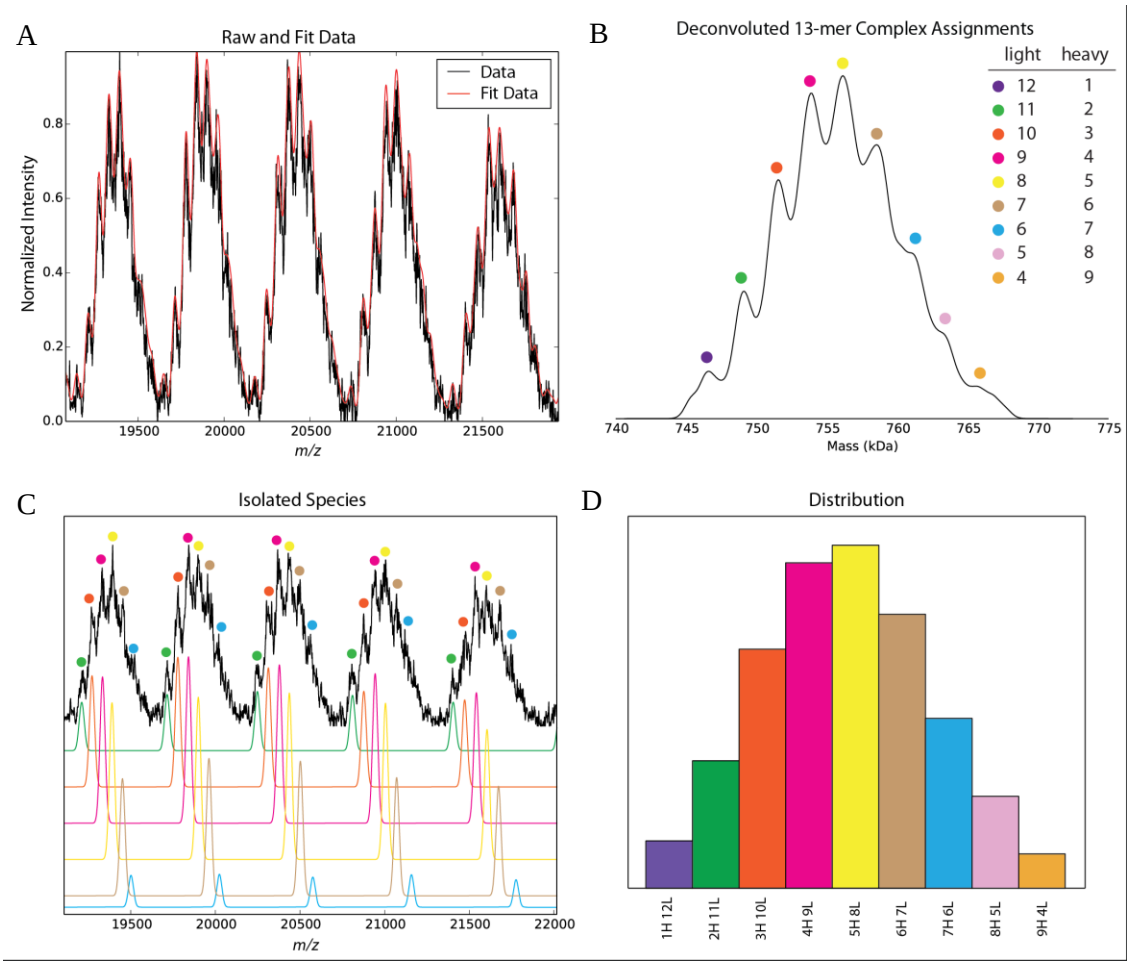
In order to experimentally demonstrate the stochastic nature of subunit mixing, we have recently begun a collaboration with the lab of Dr. Justin Benesch at the University of Oxford. Samples of wild-type GroEL grown in standard growth medium were mixed, according to the procedure outlined in A.4.2, with samples of wild-type GroEL grown in minimal media enriched with ^{13}C glucose and ^{15}N ammonium chloride. Isotopically enriched (heavy) GroEL has a subunit mass 3315 Da larger than GroEL grown under standard conditions (light). This allows for resolution of the two species using native mass spectrometry techniques, Figure A.4.3.

Figure A.4.3 – Native Mass Spectrum of Heavy and Light $GroEL_{14}$



(Top) Native mass spectrum, obtained by the Benesch lab at University of Oxford, of $GroEL_{14}$ grown under standard conditions. (Bottom) Native mass spectrum of $GroEL_{14}$ grown in minimal media isotopically enriched with ^{13}C glucose and ^{15}N ammonium chloride, resulting in "heavy" GroEL with a subunit mass 3315 Da greater than the light sample. The two samples will be mixed according to the procedure outlined in A.4.2 to produce 14mers with mixed rings that contain a distribution of heavy and light subunits. The figure was provided by the Benesch lab.

Figure A.4.4 – Mass Spectrum of Heavy and Light Mixed GroEL$_{14}$ Complexes



Native mass spectrum, obtained by the Benesch lab at University of Oxford, of a sample of 50% light and 50% heavy GroEL that has undergone subunit mixing according to the protocol in Appendix A.4.2. The mixture produces an ensemble of species with a distribution of masses. (A) The raw data (black) and the fit (red) to a series of charge packets. The fit is obtained by assigning the predicted mass of each species a population size that fits a binomial distribution. (B) The identification of each species present in a single charge packet. (C) The separation of each species contribution to the set of charge packets in (A). (D) A histogram of the relative population of each species as predicted by the fit. The distribution of masses fits to a binomial distribution but is centered around 63% light and 37% heavy. More work is needed to conclusively demonstrate that subunit mixing is stochastic. This figure was provided by the Benesch lab.

The two samples were mixed according to the subunit mixing protocol outlined in Appendix A.4.2. The resulting sample contains 14mers with a distribution of heavy and light subunits. The mass of each individual GroEL$_{14}$ particle is dependent on the relative proportion of heavy and light subunits. If the mixing of subunits is stochastic, the relative proportion of populations with

different masses will follow a simple binomial distribution. In practice, a single subunit is ejected from the $GroEL_{14}$ complex, within the mass spectrometer, in order to create 13mers, reducing mass of the complex, increasing the m/z resolution between charge packets. The ejection of a subunit is random and will not affect the predicted distribution.

To date, only one such mixing experiment has taken place, Figure A.4.4. Light and heavy GroEL 14mers were mixed in a 50/50 ratio. The resulting distribution of masses did fit a binomial distribution, but centered at a mixing ratio of 63/37 (light/heavy). The results are encouraging, as the discrepancy can be explained by a pipetting or concentration determination error. Nonetheless, more work is forthcoming to conclusively demonstrate the stochasticity of subunit mixing in urea.

APPENDIX A.5

A.5.1 Simulation Parameters

Table A.5.1 – All Constant Parameters Used in the Simulation	
k_{FENE}	20 kcal/(mol Å^2)
R_O	2.0 Å
$\epsilon_{L\,i,j}$	1 kcal/mol
ε	10 ε_o
κ^{-1}	2.4 Å
Most values were originally used in ref (4).	

Table A.5.2 – Soft Sphere Radius Values Used for Each Amino Acid							
Residue	**Rad (Å)**	**Residue**	**Rad (Å)**	**Residue**	**Rad (Å)**	**Residue**	**Rad (Å)**
Gly	0.5	Met	3.09	Asn	2.84	Glu	2.96
Ala	2.52	Phe	3.18	Gln	3.01	His	3.04
Val	2.93	Pro	2.78	Tyr	3.23	Lys	3.18
Leu	3.09	Ser	2.59	Trp	3.39	Arg	3.28
Ile	3.09	Thr	2.81	Asp	2.79	Cys	2.74
These radiuses reflect the area spherically occupied by the side chain as it explores all possible bond angles. All backbone beads were given a soft sphere radius of 3.8							

A.5.2 Native Contact Interaction Energy

As mentioned in the main body, the value of $\epsilon_{N\,i,j}$ assumed for a particular pair of residues is calculated from the Betancourt-Thirumalai (BT) potential. The BT potential is a matrix of values such that each pair of residues corresponds to one element in the matrix. It is then corrected by a factor of 0.7, an offset value empirically identified to lead to good agreement with experiment. The final form of $\epsilon_{N\,i,j}$ used was $\epsilon_{N\,i,j} = \delta_i\delta_j\left|BT_{i,j} - 0.7\right|$ where $BT_{i,j}$ is the BT potential matrix element corresponding to residues i and j, and δ_i is 0.55 (kcal/mol)$^{1/2}$ for side chain beads and 0.74 (kcal/mol)$^{1/2}$ for backbone beads. The delta values were obtained an optimization of each parameter found in ref (5).

APPENDIX A.6

1. Lay, D. *Linear Algebra and Its Applications*. Vol. 3. Boston: Pearson, 2006.

2. Ryabova, N., V. Marchenkov, N. Kotova and G. Semisotnov. "Chaperonin Groel Reassembly: An Effect of Protein Ligands and Solvent Composition." *Biomolecules* 4, no. 2 (2014): 458-73.

3. Betancourt, M. R. and D. Thirumalai. "Pair Potentials for Protein Folding: Choice of Reference States and Sensitivity of Predicted Native States to Variations in the Interaction Schemes." *Prot Sci* 8, no. 2 (1999): 361-9.

4. Hyeon, G. H. Lorimer and D. Thirumalai. "Dynamics of Allosteric Transitions in Groel." *PNAS* 103, no. 50 (2006): 18939.

5. Liu, Z., G. Reddy, E. P. O'Brien and D. Thirumalai. "Collapse Kinetics and Chevron Plots from Simulations of Denaturant-Dependent Folding of Globular Proteins." *Proc Natl Acad Sci U S A* 108, no. 19 (2011): 7787-92.

6. Ybarra, J. and P. M. Horowitz. "Inactive Groel Monomers Can Be Isolated and Reassembled to Functional Tetradecamers That Contain Few Bound Peptides." *J Biol Chem* 270, no. 39 (1995): 22962-7.

7. Ybarra, J. and P. M. Horowitz. "Refolding and Reassembly of Active Chaperonin Groel after Denaturation." *J Biol Chem* 270, no. 38 (1995): 22113-5.

DISSERTATION REFERENCES

Anfinsen, C. B. "Principles That Govern the Folding of Protein Chains." *Science* 181, no. 4096 (1973): 223-30.

Azem, A., S. Diamant, M. Kessel, C. Weiss and P. Goloubinoff. "The Protein-Folding Activity of Chaperonins Correlates with the Symmetric Groel14(Groes7)2 Heterooligomer." *Proc Natl Acad Sci U S A* 92, no. 26 (1995): 12021-5.

Azem, A., M. Kessel and P. Goloubinoff. "Characterization of a Functional Groel14(Groes7)2 Chaperonin Hetero-Oligomer." *Science* 265, no. 5172 (1994): 653-6.

Azia, A., R. Unger and A. Horovitz. "What Distinguishes Groel Substrates from Other Escherichia Coli Proteins?" *FEBS J* 279, no. 4 (2012): 543-50.

Betancourt, M. R. and D. Thirumalai. "Exploring the Kinetic Requirements for Enhancement of Protein Folding Rates in the Groel Cavity." *J Mol Biol* 287, no. 3 (1999): 627-44.

Betancourt, M. R. and D. Thirumalai. "Pair Potentials for Protein Folding: Choice of Reference States and Sensitivity of Predicted Native States to Variations in the Interaction Schemes." *Protein Sci* 8, no. 2 (1999): 361-9.

Blackman, M, J Corrie, J Croney, G Kelly, J Eccleston and D Jameson. "Structural and Biochemical Characterization of Fluorogenic Rhodamine Labeled Malarial Protease Substrate." *Biochemistry* 41, (2002).

Bochkareva, E. S., N. M. Lissin, G. C. Flynn, J. E. Rothman and A. S. Girshovich. "Positive Cooperativity in the Functioning of Molecular Chaperone Groel." *J Biol Chem* 267, no. 10 (1992): 6796-800.

Boisvert, D. C., J. Wang, Z. Otwinowski, A. L. Horwich and P. B. Sigler. "The 2.4 a Crystal Structure of the Bacterial Chaperonin Groel Complexed with Atp Gamma S." *Nat Struct Biol* 3, no. 2 (1996): 170-7.

Braig, K., P. D. Adams and A. T. Brunger. "Conformational Variability in the Refined Structure of the Chaperonin Groel at 2.8 a Resolution." *Nat Struct Biol* 2, no. 12 (1995): 1083-94.

Braig, K., Z. Otwinowski, R. Hegde, D. C. Boisvert, A. Joachimiak, A. L. Horwich and P. B. Sigler. "The Crystal Structure of the Bacterial Chaperonin Groel at 2.8 A." *Nature* 371, no. 6498 (1994): 578-86.

Brune, M., J. L. Hunter, J. E. Corrie and M. R. Webb. "Direct, Real-Time Measurement of Rapid Inorganic Phosphate Release Using a Novel Fluorescent Probe and Its Application to Actomyosin Subfragment 1 Atpase." *Biochemistry* 33, no. 27 (1994): 8262-71.

Burston, S. G., N. A. Ranson and A. R. Clarke. "The Origins and Consequences of Asymmetry in the Chaperonin Reaction Cycle." *J Mol Biol* 249, no. 1 (1995): 138-52.

Chaudhry, Charu, Arthur L. Horwich, Axel T. Brunger and Paul D. Adams. "Exploring the Structural Dynamics of the E.Coli Chaperonin Groel Using Translation-Libration-Screw Crystallographic Refinement of Intermediate States." *Journal of Molecular Biology* 342,

no. 1 (2004): 229-245.

Chen, L. and P. B. Sigler. "The Crystal Structure of a Groel/Peptide Complex: Plasticity as a Basis for Substrate Diversity." *Cell* 99, no. 7 (1999): 757-68.

Cleland, W. W. "Dithiothreitol, a New Protective Reagent for Sh Groups." *Biochemistry* 3, (1964): 480-2.

Corsepius, N. C. and G. H. Lorimer. "Measuring How Much Work the Chaperone Groel Can Do." *Proc Natl Acad Sci U S A* 110, no. 27 (2013): E2451-9.

Cui, Q. and M. Karplus. "Allostery and Cooperativity Revisited." *Protein Sci* 17, no. 8 (2008): 1295-307.

Danziger, O., D. Rivenzon-Segal, S. G. Wolf and A. Horovitz. "Conversion of the Allosteric Transition of Groel from Concerted to Sequential by the Single Mutation Asp-155 -> Ala." *Proc Natl Acad Sci U S A* 100, no. 24 (2003): 13797-802.

DeVore, J. *Probability and Statistics for Engineers and Scientists*. Vol. 7. Belmont: Thomson, 2008.

Dima, R. I. and H. Joshi. "Probing the Origin of Tubulin Rigidity with Molecular Simulations." *Proc Natl Acad Sci U S A* 105, no. 41 (2008): 15743-8.

Edmundson, A. B., K. R. Ely and J. N. Herron. "A Search for Site-Filling Ligands in the Mcg Bence-Jones Dimer: Crystal Binding Studies of Fluorescent Compounds." *Mol Immunol* 21, no. 7 (1984): 561-76.

Ellis, J. "Proteins as Molecular Chaperones." *Nature* 328, no. 6129 (1987): 378-9.

Engel, A., M. K. Hayer-Hartl, K. N. Goldie, G. Pfeifer, R. Hegerl, S. Muller, A. C. da Silva, W. Baumeister and F. U. Hartl. "Functional Significance of Symmetrical Versus Asymmetrical Groel-Groes Chaperonin Complexes." *Science* 269, no. 5225 (1995): 832-6.

Fei, X., D. Yang, N. LaRonde-LeBlanc and G. H. Lorimer. "Crystal Structure of a Groel-Adp Complex in the Relaxed Allosteric State at 2.7 a Resolution." *Proc Natl Acad Sci U S A* 110, no. 32 (2013): E2958-66.

Fenton, W. A., Y. Kashi, K. Furtak and A. L. Horwich. "Residues in Chaperonin Groel Required for Polypeptide Binding and Release." *Nature* 371, no. 6498 (1994): 614-9.

Frank, G. A., M. Goomanovsky, A. Davidi, G. Ziv, A. Horovitz and G. Haran. "Out-of-Equilibrium Conformational Cycling of Groel under Saturating Atp Concentrations." *Proc Natl Acad Sci U S A* 107, no. 14 (2010): 6270-4.

Friedman, D. I., E. R. Olson, C. Georgopoulos, K. Tilly, I. Herskowitz and F. Banuett. "Interactions of Bacteriophage and Host Macromolecules in the Growth of Bacteriophage Lambda." *Microbiol Rev* 48, no. 4 (1984): 299-325.

Gorovits, B. M., J. W. Seale and P. M. Horowitz. "Residual Structure in Urea-Denatured

Chaperonin Groel." *Biochemistry* 34, no. 42 (1995): 13928-33.

Gorovits, B. M., J. Ybarra and P. M. Horowitz. "Atp Hydrolysis Is Critical for Induction of Conformational Changes in Groel That Expose Hydrophobic Surfaces." *J Biol Chem* 272, no. 11 (1997): 6842-5.

Grason, J. P., J. S. Gresham and G. H. Lorimer. "Setting the Chaperonin Timer: A Two-Stroke, Two-Speed, Protein Machine." *Proc Natl Acad Sci U S A* 105, no. 45 (2008): 17339-44.

Grason, J. P., J. S. Gresham, L. Widjaja, S. C. Wehri and G. H. Lorimer. "Setting the Chaperonin Timer: The Effects of K+ and Substrate Protein on Atp Hydrolysis." *Proceedings of the National Academy of Sciences* 105, no. 45 (2008): 17334-17338.

Gray, T. E. and A. R. Fersht. "Cooperativity in Atp Hydrolysis by Groel Is Increased by Groes." *FEBS Lett* 292, no. 1-2 (1991): 254-8.

Gresham, J.S. "Allostery and Groel: Exploring the Tenets of Nested Cooperativity." University of Maryland, 2004.

Hamman, B. D., A. V. Oleinikov, G. G. Jokhadze, D. E. Bochkariov, R. R. Traut and D. M. Jameson. "Tetramethylrhodamine Dimer Formation as a Spectroscopic Probe of the Conformation of Escherichia Coli Ribosomal Protein L7/L12 Dimers." *J Biol Chem* 271, no. 13 (1996): 7568-73.

Harris, J. R., A. Pluckthun and R. Zahn. "Transmission Electron Microscopy of Groel, Groes, and the Symmetrical Groel/Es Complex." *J Struct Biol* 112, no. 3 (1994): 216-30.

Hemmingsen, S. M., C. Woolford, S. M. van der Vies, K. Tilly, D. T. Dennis, C. P. Georgopoulos, R. W. Hendrix and R. J. Ellis. "Homologous Plant and Bacterial Proteins Chaperone Oligomeric Protein Assembly." *Nature* 333, no. 6171 (1988): 330-4.

Hendrix, R. W. "Purification and Properties of Groe, a Host Protein Involved in Bacteriophage Assembly." *J Mol Biol* 129, no. 3 (1979): 375-92.

Hohn, T., B. Hohn, A. Engel, M. Wurtz and P. R. Smith. "Isolation and Characterization of the Host Protein Groe Involved in Bacteriophage Lambda Assembly." *Journal of Molecular Biology* 129, no. 3 (1979): 359-73.

Horovitz, A., Y. Fridmann, G. Kafri and O. Yifrach. "Review: Allostery in Chaperonins." *J Struct Biol* 135, no. 2 (2001): 104-14.

Horowitz, G. H. Lorimer and Ybarra. "Groes in the Asymmetric Groel Groes Complex Exchanges Via an Associative Mechanism." *PNAS* 96, no. 6 (1999): 2682.

Horwich, Arthur L., Adrian C. Apetri and Wayne A. Fenton. "The Groel/Groes Cis Cavity as a Passive Anti-Aggregation Device." *FEBS Letters* 583, no. 16 (2009): 2654-2662.

Horwich, A. L., G. W. Farr and W. A. Fenton. "Groel-Groes-Mediated Protein Folding." *Chem Rev* 106, no. 5 (2006): 1917-30.

Horwich, A. L. and W. A. Fenton. "Chaperonin-Mediated Protein Folding: Using a Central

Cavity to Kinetically Assist Polypeptide Chain Folding." *Q Rev Biophys* 42, no. 2 (2009): 83-116.

Hunt, J. F., A. J. Weaver, S. J. Landry, L. Gierasch and J. Deisenhofer. "The Crystal Structure of the Groes Co-Chaperonin at 2.8 a Resolution." *Nature* 379, no. 6560 (1996): 37-45.

Hyeon, G. H. Lorimer and D. Thirumalai. "Dynamics of Allosteric Transitions in Groel." *PNAS* 103, no. 50 (2006): 18939.

Hyeon, C., R. I. Dima and D. Thirumalai. "Pathways and Kinetic Barriers in Mechanical Unfolding and Refolding of Rna and Proteins." *Structure* 14, no. 11 (2006): 1633-45.

Hyeon, C., P. A. Jennings, J. A. Adams and J. N. Onuchic. "Ligand-Induced Global Transitions in the Catalytic Domain of Protein Kinase A." *Proc Natl Acad Sci U S A* 106, no. 9 (2009): 3023-8.

Hyeon, C. and D. Thirumalai. "Mechanical Unfolding of Rna: From Hairpins to Structures with Internal Multiloops." *Biophys J* 92, no. 3 (2007): 731-43.

Ilich, P., P.K. Mishra, S. Macura and T.P. Burghardt. "Direct Observation of Rhodamine Dimer Structure in Water." *Spectrochimica acta. Part A, Molecular Spectroscopy* 52, no. 10 (1996): 1323-1330.

Kasha, M. "Energy Transfer Mechanisms and the Molecular Exciton Model for Molecular Aggregates." *Radiat Res* 20, (1963): 55-70.

Kipnis, Y., N. Papo, G. Haran and A. Horovitz. "Concerted Atp-Induced Allosteric Transitions in Groel Facilitate Release of Protein Substrate Domains in an All-or-None Manner." *Proceedings of the National Academy of Sciences* 104, no. 9 (2007): 3119-3124.

Koike-Takeshita, A., M. Yoshida and H. Taguchi. "Revisiting the Groel-Groes Reaction Cycle Via the Symmetric Intermediate Implied by Novel Aspects of the Groel(D398a) Mutant." *J Biol Chem* 283, no. 35 (2008): 23774-81.

Landry, S. J. and L. M. Gierasch. "Polypeptide Interactions with Molecular Chaperones and Their Relationship to in Vivo Protein Folding." *Annu Rev Biophys Biomol Struct* 23, (1994): 645-69.

Landry, S. J., J. Zeilstra-Ryalls, O. Fayet, C. Georgopoulos and L. M. Gierasch. "Characterization of a Functionally Important Mobile Domain of Groes." *Nature* 364, no. 6434 (1993): 255-8.

Langer, T., G. Pfeifer, J. Martin, W. Baumeister and F. U. Hartl. "Chaperonin-Mediated Protein Folding: Groes Binds to One End of the Groel Cylinder, Which Accommodates the Protein Substrate within Its Central Cavity." *EMBO J* 11, no. 13 (1992): 4757-65.

Lay, D. *Linear Algebra and Its Applications*. Vol. 3. Boston: Pearson, 2006.

Lin, J. C. and D. Thirumalai. "Relative Stability of Helices Determines the Folding Landscape of Adenine Riboswitch Aptamers." *J Am Chem Soc* 130, no. 43 (2008): 14080-1.

Lin, Zong, Damian Madan and Hays S. Rye. "Groel Stimulates Protein Folding through Forced Unfolding." *Nature Structural & Molecular Biology* 15, no. 3 (2008): 303-311.

Lin, Zong and Hays S. Rye. "Groel-Mediated Protein Folding: Making the Impossible, Possible." *Critical Reviews in Biochemistry and Molecular Biology* 41, no. 4 (2006): 211-239.

Lin, Z., F. P. Schwartz and E. Eisenstein. "The Hydrophobic Nature of Groel-Substrate Binding." *J Biol Chem* 270, no. 3 (1995): 1011-4.

Liu, Z., G. Reddy, E. P. O'Brien and D. Thirumalai. "Collapse Kinetics and Chevron Plots from Simulations of Denaturant-Dependent Folding of Globular Proteins." *Proc Natl Acad Sci U S A* 108, no. 19 (2011): 7787-92.

Llorca, O., J. L. Carrascosa and J. M. Valpuesta. "Biochemical Characterization of Symmetric Groel-Groes Complexes. Evidence for a Role in Protein Folding." *J Biol Chem* 271, no. 1 (1996): 68-76.

Llorca, O., S. Marco, J. L. Carrascosa and J. M. Valpuesta. "Symmetric Groel-Groes Complexes Can Contain Substrate Simultaneously in Both Groel Rings." *FEBS Lett* 405, no. 2 (1997): 195-9.

Lorimer, G. "Protein Folding. Folding with a Two-Stroke Motor." *Nature* 388, no. 6644 (1997): 720-1, 723.

Lorimer, G. H. "A Personal Account of Chaperonin History." *Plant Physiology* 125, no. 1 (2001): 38-41.

Ma, J., P. B. Sigler, Z. Xu and M. Karplus. "A Dynamic Model for the Allosteric Mechanism of Groel." *J Mol Biol* 302, no. 2 (2000): 303-13.

Madan, D., Z. Lin and H. S. Rye. "Triggering Protein Folding within the Groel-Groes Complex." *Journal of Biological Chemistry* 283, no. 46 (2008): 32003-32013.

Mendoza, J. A., B. Demeler and P. M. Horowitz. "Alteration of the Quaternary Structure of Cpn60 Modulates Chaperonin-Assisted Folding. Implications for the Mechanism of Chaperonin Action." *J Biol Chem* 269, no. 4 (1994): 2447-51.

Mendoza, J. A., J. L. Martinez and P. M. Horowitz. "Tetradecameric Chaperonin 60 Can Be Assembled in Vitro from Monomers in a Process That Is Atp Independent." *Biochim Biophys Acta* 1247, no. 2 (1995): 209-14.

Mickler, M., R. I. Dima, H. Dietz, C. Hyeon, D. Thirumalai and M. Rief. "Revealing the Bifurcation in the Unfolding Pathways of Gfp by Using Single-Molecule Experiments and Simulations." *Proc Natl Acad Sci U S A* 104, no. 51 (2007): 20268-73.

Motojima, Charu Chaudhry, W. A. Fenton, G. W. Farr and A. L. Horwich. "Substrate Polypeptide Presents a Load on the Apical Domains of the Chaperonin Groel." *PNAS* 101, no. 42 (2002): 15005.

Murai, N., Y. Makino and M. Yoshida. "Groel Locked in a Closed Conformation by an Interdomain Cross-Link Can Bind Atp and Polypeptide but Cannot Process Further

Reaction Steps." *J Biol Chem* 271, no. 45 (1996): 28229-34.

Okoh, M. P., J. L. Hunter, J. E. Corrie and M. R. Webb. "A Biosensor for Inorganic Phosphate Using a Rhodamine-Labeled Phosphate Binding Protein." *Biochemistry* 45, no. 49 (2006): 14764-71.

Packard, Toptygin, Komoriya and Brand. "Profluorescent Protease Substrates Intramolecular Dimers Described by the Exciton Model." *Proceedings of the National Academy of Sciences* 93, (1996): 11640.

Papo, N., Y. Kipnis, G. Haran and A. Horovitz. "Concerted Release of Substrate Domains from Groel by Atp Is Demonstrated with Fret." *J Mol Biol* 380, no. 4 (2008): 717-25.

Ranson, Neil A., Daniel K. Clare, George W. Farr, David Houldershaw, Arthur L. Horwich and Helen R. Saibil. "Allosteric Signaling of Atp Hydrolysis in Groel–Groes Complexes." *Nature Structural & Molecular Biology* 13, no. 2 (2006): 147-152.

Ranson, N. A., N. J. Dunster, S. G. Burston and A. R. Clarke. "Chaperonins Can Catalyse the Reversal of Early Aggregation Steps When a Protein Misfolds." *J Mol Biol* 250, no. 5 (1995): 581-6.

Ranson, N. A., G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich and H. R. Saibil. "Atp-Bound States of Groel Captured by Cryo-Electron Microscopy." *Cell* 107, no. 7 (2001): 869-79.

Roseman, A. M., S. Chen, H. White, K. Braig and H. R. Saibil. "The Chaperonin Atpase Cycle: Mechanism of Allosteric Switching and Movements of Substrate-Binding Domains in Groel." *Cell* 87, no. 2 (1996): 241-51.

Ryabova, N., V. Marchenkov, N. Kotova and G. Semisotnov. "Chaperonin Groel Reassembly: An Effect of Protein Ligands and Solvent Composition." *Biomolecules* 4, no. 2 (2014): 458-73.

Rye, H. S., S. G. Burston, W. A. Fenton, J. M. Beechem, Z. Xu, P. B. Sigler and A. L. Horwich. "Distinct Actions of Cis and Trans Atp within the Double Ring of the Chaperonin Groel." *Nature* 388, no. 6644 (1997): 792-8.

Saibil, H., Z. Dong, S. Wood and A. auf der Mauer. "Binding of Chaperonins." *Nature* 353, no. 6339 (1991): 25-6.

Saibil, H. R., W. A. Fenton, D. K. Clare and A. L. Horwich. "Structure and Allostery of the Chaperonin Groel." *J Mol Biol* 425, no. 9 (2013): 1476-87.

Sameshima, T., R. Iizuka, T. Ueno and T. Funatsu. "Denatured Proteins Facilitate the Formation of the Football-Shaped Groel-(Groes)2 Complex." *Biochem J* 427, no. 2 (2010): 247-54.

Schlick, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Vol. 21. 2 ed. New York: Springer, 2010.

Schmidt, M., K. Rutkat, R. Rachel, G. Pfeifer, R. Jaenicke, P. Viitanen, G. Lorimer and J. Buchner. "Symmetric Complexes of Groe Chaperonins as Part of the Functional Cycle."

Science 265, no. 5172 (1994): 656-9.

Shiseki, K., N. Murai, F. Motojima, T. Hisabori, M. Yoshida and H. Taguchi. "Synchronized Domain-Opening Motion of Groel Is Essential for Communication between the Two Rings." *J Biol Chem* 276, no. 14 (2001): 11335-8.

Sorkine, O. "As-Rigid-as-Possible Surface Modeling." *AMC International Conference Proceeding Series* 257, (2007): 109-116.

Spiegel, M. *Schaum's Outline of Theory and Problems of Laplace Transforms* Schaum's Outline. New York: McGraw-Hill, 1965.

Stan, G., B. R. Brooks, G. H. Lorimer and D. Thirumalai. "Identifying Natural Substrates for Chaperonins Using a Sequence-Based Approach." *Protein Sci* 14, no. 1 (2005): 193-201.

Stan, G., B. R. Brooks, G. H. Lorimer and D. Thirumalai. "Residues in Substrate Proteins That Interact with Groel in the Capture Process Are Buried in the Native State." *Proc Natl Acad Sci U S A* 103, no. 12 (2006): 4433-8.

Takano, T. and T. Kakefuda. "Involvement of a Bacterial Factor in Morphogenesis of Bacteriophage Capsid." *Nat New Biol* 239, no. 89 (1972): 34-7.

Takei, Y., R. Iizuka, T. Ueno and T. Funatsu. "Single-Molecule Observation of Protein Folding in Symmetric Groel-(Groes)2 Complexes." *J Biol Chem* 287, no. 49 (2012): 41118-25.

Tehver, R., J. Chen and D. Thirumalai. "Allostery Wiring Diagrams in the Transitions That Drive the Groel Reaction Cycle." *J Mol Biol* 387, no. 2 (2009): 390-406.

Thirumalai, D. and G. H. Lorimer. "Chaperonin-Mediated Protein Folding." *Annu Rev Biophys Biomol Struct* 30, (2001): 245-69.

Todd, M. J., G. H. Lorimer and D. Thirumalai. "Chaperonin-Facilitated Protein Folding: Optimization of Rate and Yield by an Iterative Annealing Mechanism." *Proc Natl Acad Sci U S A* 93, no. 9 (1996): 4030-5.

Todd, M. J., P. V. Viitanen and G. H. Lorimer. "Dynamics of the Chaperonin Atpase Cycle: Implications for Facilitated Protein Folding." *Science* 265, no. 5172 (1994): 659-66.

Toptygin, D., B. Packard and L. Brand. "Resolution of Absorption Spectra of Rhodamine 6g Aggregates in Aqueous Solution Using the Law of Mass Action." *Chemical Physics Letters* 5, no. 227 (1997): 430-435.

Tossavainen, T. "On the Zeros of Finite Sums of Exponential Functions." *Mathematical Society Gazette* 1, no. 33 (2006): 47-50.

Tyagi, Navneet K., Wayne A. Fenton and Arthur L. Horwich. "Atp-Triggered Adp Release from the Asymmetric Chaperonin Groel/Groes/Adp7 Is Not the Rate-Limiting Step of the Groel/Groes Reaction Cycle." *FEBS Letters* 584, no. 5 (2010): 951-953.

Ueno, T., H. Taguchi, H. Tadakuma, M. Yoshida and T. Funatsu. "Groel Mediates Protein Folding with a Two Successive Timer Mechanism." *Mol Cell* 14, no. 4 (2004): 423-34.

Viitanen, P. V., A. A. Gatenby and G. H. Lorimer. "Purified Chaperonin 60 (Groel) Interacts with the Nonnative States of a Multitude of Escherichia Coli Proteins." *Protein Sci* 1, no. 3 (1992): 363-9.

Viitanen, P. V., T. H. Lubben, J. Reed, P. Goloubinoff, D. P. O'Keefe and G. H. Lorimer. "Chaperonin-Facilitated Refolding of Ribulosebisphosphate Carboxylase and Atp Hydrolysis by Chaperonin 60 (Groel) Are K+ Dependent." *Biochemistry* 29, no. 24 (1990): 5665-71.

Weibe, A. "Global Sensitivity Analysis of Ordinary Differential Equations." Free University of Berlin, 2009.

Weissman, J. S., C. M. Hohl, O. Kovalenko, Y. Kashi, S. Chen, K. Braig, H. R. Saibil, W. A. Fenton and A. L. Horwich. "Mechanism of Groel Action: Productive Release of Polypeptide from a Sequestered Position under Groes." *Cell* 83, no. 4 (1995): 577-87.

Xu, Z., A. L. Horwich and P. B. Sigler. "The Crystal Structure of the Asymmetric Groel-Groes-(Adp)7 Chaperonin Complex." *Nature* 388, no. 6644 (1997): 741-50.

Xu, Z. and P. B. Sigler. "Groel/Groes: Structure and Function of a Two-Stroke Folding Machine." *J Struct Biol* 124, no. 2-3 (1998): 129-41.

Yang, D., X. Ye and G. H. Lorimer. "Symmetric Groel:Groes2 Complexes Are the Protein-Folding Functional Form of the Chaperonin Nanomachine." *Proc Natl Acad Sci U S A* 110, no. 46 (2013): E4298-305.

Yang, Y. "Site-Directed Mutagenesis of Groel: Developing a System for Monitoring Allosteric Movements by Fluorescence Resonance Energy Transfer (Fret)." University of Maryland, 2006.

Ybarra, J. and P. M. Horowitz. "Inactive Groel Monomers Can Be Isolated and Reassembled to Functional Tetradecamers That Contain Few Bound Peptides." *J Biol Chem* 270, no. 39 (1995): 22962-7.

Ybarra, J. and P. M. Horowitz. "Refolding and Reassembly of Active Chaperonin Groel after Denaturation." *J Biol Chem* 270, no. 38 (1995): 22113-5.

Ye, X. and G. H. Lorimer. "Substrate Protein Switches Groe Chaperonins from Asymmetric to Symmetric Cycling by Catalyzing Nucleotide Exchange." *Proc Natl Acad Sci U S A* 110, no. 46 (2013): E4289-97.

Yifrach, O. and A. Horovitz. "Two Lines of Allosteric Communication in the Oligomeric Chaperonin Groel Are Revealed by the Single Mutation Arg196-->Ala." *J Mol Biol* 243, no. 3 (1994): 397-401.

Yifrach, Ofer and A. Horovitz. "Nested Cooperativity in the Atpase Activity of the Oligomeric Chaperonin Groel." *Biochemistry* 34, no. 16 (1995): 5303.

Yifrach, O. and A. Horovitz. "Allosteric Control by Atp of Non-Folded Protein Binding to Groel." *J Mol Biol* 255, no. 3 (1996): 356-61.

Yifrach, O. and A. Horovitz. "Transient Kinetic Analysis of Adenosine 5'-Triphosphate Binding-Induced Conformational Changes in the Allosteric Chaperonin Groel." *Biochemistry* 37, no. 20 (1998): 7083-8.